JOURNAL OF BACTERIOLOGY, Dec. 2004, p. 8074–8082 Vol. 186, No. 23
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.23.8074–8082.2004

Crystal Structure of the PdxY Protein from *Escherichia coli*

Martin K. Safo,[1,2]* Faik N. Musayev,[1,2] Sharyn Hunt,[2,3] Martino L. di Salvo,[4] Neel Scarsdale,[2,3] and Verne Schirch[2,3]

Department of Medicinal Chemistry,[1] Department of Biochemistry,[3] and Institute for Structural Biology and Drug Discovery,[2] Virginia Commonwealth University, Richmond, Virginia, and Dipartimento di Scienze Biochimiche "A. Rossi Fanelli," Università La Sapienza, Rome, Italy[4]

Received 1 June 2004/Accepted 25 August 2004

The crystal structure of *Escherichia coli* PdxY, the protein product of the *pdxY* gene, has been determined to a 2.2-Å resolution. PdxY is a member of the ribokinase superfamily of enzymes and has sequence homology with pyridoxal kinases that phosphorylate pyridoxal at the C-5′ hydroxyl. The protein is a homodimer with an active site on each monomer composed of residues that come exclusively from each respective subunit. The active site is filled with a density that fits that of pyridoxal. In monomer A, the ligand appears to be covalently attached to Cys122 as a thiohemiacetal, while in monomer B it is not covalently attached but appears to be partially present as pyridoxal 5′-phosphate. The presence of pyridoxal phosphate and pyridoxal as ligands was confirmed by the activation of aposerine hydroxymethyltransferase after release of the ligand by the denaturation of PdxY. The ligand, which appears to be covalently attached to Cys122, does not dissociate after denaturation of the protein. A detailed comparison (of functional properties, sequence homology, active site and ATP-binding-site residues, and active site flap types) of PdxY with other pyridoxal kinases as well as the ribokinase superfamily in general suggested that PdxY is a member of a new subclass of the ribokinase superfamily. The structure of PdxY also permitted an interpretation of work that was previously published about this enzyme.

Pyridoxal phosphate (PLP) serves as a cofactor for many enzymes that are involved in amino acid and sugar metabolism. In many bacteria and plants, PLP is synthesized by a de novo pathway, but most cells rely on a nutritional source of vitamin B_6, i.e., pyridoxine (PN), pyridoxal (PL), or pyridoxamine (PM) (18). All cells, however, have a salvage pathway for reutilizing the PLP that is liberated during protein turnover (21). The salvage pathway involves an ATP-dependent pyridoxal kinase that phosphorylates PL, PN, and PM. The product of PN and PM phosphorylation is converted to PLP by pyridoxine 5-phosphate oxidase, which is also expressed in most cells. Because those PL kinases that have been purified and studied show activity toward PN and PM in addition to PL, they are often referred to as PL/PN/PM kinases. For the sake of brevity, we will refer to these enzymes as PL kinases with the understanding that many of them exhibit considerable activity toward PN and PM.

PL kinases have been purified from bacterial, plant, and mammalian sources, and recently the crystal structure of the sheep brain kinase was solved (9, 10). The sheep enzyme was found to be a member of the ribokinase superfamily, but with some unique features in its mechanism for protecting ATP hydrolysis in the absence of the cosubstrate PL (9). Evidence suggests that most organisms contain a single PL kinase. However, some studies showed that *Escherichia coli* mutants that were blocked in the de novo synthetic pathway and had an inactivated *pdxK* gene, which codes for PL kinase 1, were still able to grow on PL (20). A search for a gene that expressed a protein that had PL kinase activity led to the discovery of the *pdxY* gene (20). The protein expressed by this gene was shown to have some PL kinase activity, but at a highly reduced level compared to the protein expressed by the *pdxK* gene (5, 20). It was postulated that the protein expressed by the *pdxY* gene was a kinase in another metabolic pathway but that it had enough activity with PL as a substrate to support growth in the absence of the de novo biosynthetic pathway and PL kinase 1. A structure for the protein encoded by the *pdxY* gene has been deposited in the Protein Data Bank (PDB) under the name pyridoxamine kinase (1VI9), but to our knowledge its structure and properties have not been described. The function of the PdxY protein in *E. coli* has not been established, as the purified enzyme shows little activity for phosphorylating PL, PN, and PM (5).

Interest in PL kinase centers on its key role in brain metabolism because the synthesis of most neurotransmitters involves PLP-dependent enzymes. Disruption of the salvage pathway or a lack of nutritional intake of vitamin B_6 results in neurological disorders (8). PL kinase activity appears to be a key site for these disorders, and several drugs are available to modulate this activity. Understanding the structure and function of PL kinase is important for advancing our knowledge of the drugs used in these therapies.

We have cloned and expressed the proteins from the *pdxK* and *pdxY* genes in *E. coli* (5). Here we report the structure of the PdxY protein. Unlike the structure that has been deposited in the PDB, our PdxY structure contains significant electron density at the active sites of both subunits, where the putative substrate PL would bind. In one subunit, this density appears to be covalently attached to Cys122. The finding of PL and PLP bound tightly at the active site supports the view that the PdxY protein plays some role in vitamin B_6 metabolism. The structure of PdxY was compared to the structures of sheep brain PL kinase and other members of the ribokinase superfamily of enzymes. The structures of the proteins, as well as their functional properties, sequence homology, active site and ATP-

* Corresponding author. Mailing address: Department of Medicinal Chemistry and Institute for Structural Biology and Drug Discovery, 800 E. Leigh St., Virginia Commonwealth University, Richmond, VA 23219. Phone: (804) 828-7291. Fax: (804) 827-3664. E-mail: msafo@mail2.vcu.edu.

binding-site residues, and active site flap types, suggest that PdxY and similar proteins in other prokaryotes may form a new subclass in the ribokinase superfamily.

MATERIALS AND METHODS

Materials. All buffers and reagents used were of the highest purity available. The enzymes that were used to assay for PL, PLP, and serine hydroxymethyltransferase activity were purified as described previously (7). The *pdxY* gene (accession no. NC000913) was cloned into a pET22 (+) vector and transformed into *E. coli* HMS174(DE3) competent cells as described previously (5).

Expression and purification of PdxY. The expression, purification, and determination of kinetic constants of PdxY were recently described (5). Briefly, *E. coli* HMS174(DE3) containing a plasmid with a *pdxY* insert was grown in 6 liter of 1.5× Luria broth in a New Brunswick BioFlo 110 fermentor at 37°C. Isopropyl-β-D-thiogalactopyranoside was added to 0.5 mM after the optical density at 600 nm reached 1.5. The temperature was then reduced to 30°C, and incubation was continued for another 6 to 7 h. The cells were harvested and then ruptured by an osmotic shock. Streptomycin sulfate was added to a final concentration of 10 g/liter to remove excess nucleic acids. After an ammonium sulfate fractionation, the enzyme was purified on phenyl Sepharose and trimethylaminoethyl columns.

Denaturation and chromophore identification. The enzyme (2 to 6 mg) was denatured under each of the following conditions: (i) 0.1 M NaOH; (ii) a boiling water bath for 10 min followed by centrifugation; and (iii) 0.1 M citrate buffer, pH 3.3, followed by a boiling water bath for 10 min, neutralization to pH 7.0 with NaOH, and centrifugation. The supernatants from these denaturation steps were then incubated with 0.5 mg of catalytically inactive *E. coli* aposerine hydroxymethyltransferase (eSHMT) in either the absence or presence of 1 mM MgATP. A supernatant from another PdxY extract was incubated with 0.5 mg of apo-eSHMT, 1 mM MgATP, and 0.35 mg of ePL kinase 1 at 37°C for 30 min. The eSHMT was then checked for catalytic activity as follows. The respective incubation samples (400 μl) were placed in a thermostated 37°C cuvette with 325 μl of a mixture containing 20 mM potassium phosphate (pH 7.3), 5 mM 2-mercaptoethanol, 0.24 mg NADP, 30 μmol of L-serine, 10 nmol of tetrahydrofolate, and 8 μmol of methylenetetrahydrofolate dehydrogenase. The activity was determined by monitoring the A_{340} (formation of NADPH) over the course of 1 min. The initial rate of this assay was compared to a standard curve prepared by substituting increasing amounts of PLP in place of the supernatants from the PdxY denaturation steps.

Crystallization and X-ray data collection and processing. Crystallization experiments were performed via a hanging-drop vapor diffusion technique at room temperature. The hanging drops were formed by mixing equal volumes of the protein solution (28 mg of enzyme/ml in 20 mM potassium phosphate, pH 7.0) and the precipitating solution. Crystals with thin, plate-like morphology grew from 1.2 to 1.6 M $(NH_4)_2SO_4$ solutions buffered with 0.1 M sodium 2-[*N*-morpholino]ethanesulfonate, pH 6.0.

The crystals belonged to the $P2_1$ space group, with typical unit cell dimensions as follows: a = 63.13 Å, b = 67.38 Å, c = 73.59 Å, and β = 93.67°. Initially, a 2.6-Å data set was obtained at 100 K by use of an in-house X-Stream cryogenic crystal cooler system (Molecular Structure Corporation [MSC], The Woodlands, Tex.), an R-Axis II image plate detector equipped with OSMIC confocal mirrors, and a Rigaku RU-200 X-ray generator operating at 50 kV and 100 mA. The data set was processed with the MSC Biotex software program. During structure refinement, a second 2.22-Å data set collected at the National Synchrotron Light Source facility at 100 K became available. The intensity data were processed and scaled with Denzo/Scalepack (14). For data collection, crystals were soaked for a few seconds in a solution containing 2.1 M $(NH_4)_2SO_4$, 17% glycerol, and 0.1 M sodium 2-[*N*-morpholino]ethanesulfonate (pH 6.2) and then transferred to the same solution, but with 29% glycerol. Details of the data collection statistics from the National Synchrotron Light Source are given in Table 1.

Solution of the PdxY structure. The crystal structure of PdxY was solved by the molecular replacement method of the AMoRe program (13). Based on the solvent content of the unit cell (~51%), we expected one dimer of PdxY per asymmetric unit, and therefore the dimeric structure of sheep PL kinase (PDB code 1LHR), with all solvent molecules omitted, was used as the search model. The cross-rotation function was calculated by using normalized structure factors for data between 8.0 and 4.0 Å. A translation function followed by rigid body refinement for the space group $P2_1$ resulted in a solution with a correlation coefficient of 0.21 and an R factor of 52.4%. The next peak had a corresponding correlation coefficient of 0.10 and an R factor of 55.2%. The use of a polyalanine model for molecular replacement also yielded similar statistics.

Refinement and model building of the PdxY structure. The following corrections were made to the sheep PL kinase molecular replacement solution model before refinement. All nonglycine and nonalanine amino acids were mutated to alanine. All insertions, plus two residues before and after each insertion, were deleted. Two residues before and after each gap were also deleted. Finally, a whole section of residues from Ala191 to Tyr198 was deleted because of close contacts with a symmetry-related molecule. The refinement was performed with the CNS program (1), with a bulk solvent correction.

TABLE 1. Data collection and refinement statistics for PdxY

Parameter	Value[a]
Data collection statistics	
Space group	$P2_1$
Unit cell parameters (Å)	63.13, 67.38, 73.59; 93.67°
Resolution limits (Å)	50.0–2.21 (2.29)
No. of measurements	67,994
No. of unique reflections	28,233
Redundancy	2.4 (1.9)
$I/\Sigma I$	11.9 (4.0)
Completeness (%)	90.3 (89.1)
R_{merge} (%)[b]	12.8 (37.7)
Structure refinement statistics	
Resolution limit (Å)	38.20–2.22 (2.36–2.22)
No. of reflections (all data)	27,678 (4,372)
R factor (%)	17.7 (22.7)
R-free value (%)	24.6 (31.7)
RMSD from standard geometry	
Bond length (Å)	0.011
Bond angle (°)	1.6
Average B values (Å^2)	
Protein/water atoms	25.6/33.7
PL/sulfate atoms	24.8/57.3
Ramachandran plot (%)	
In most favored/additional	90.0/10.0
Estimated coordinate errors (Luzzati plot)	
R factor	0.21
R-free value	0.34

[a] Numbers in parentheses refer to the outermost resolution bin.
[b] $R_{merge} = \sigma(\langle I \rangle - I)/\Sigma I$.

The starting model was subjected to rigid body, positional, simulated annealing, and individual B-factor refinements, using all 2.6-Å crystallographic data, to an R factor and an R-free value of 37.1 and 45.7%, respectively. The model resulted in a clear density for some of the side chains, which were then modeled into the map. Further refinements with intermittent model corrections resulted in an R factor and an R-free value of 35.0 and 42.9%, respectively, with about 55% of the side chain residues built into the model. At this stage, we obtained a new data set to 2.22 Å from the National Synchrotron Light Source, which was then used for all subsequent refinements. The resulting map at 2.4 Å, after a round of rigid body, positional, simulated annealing, and B-factor refinements (R factor = 34.0% and R-free value = 40.5%), showed a significantly improved electron density map that was easily interpretable. An almost complete model with side chains was then built with ARP/wARP, version 6.0 (16, 19), at 2.3 Å. The only missing segments were residues 117 to 121 and two residues each at the C-terminal ends of the two chains. The model was subsequently refined against the 2.22-Å data set with all of the measured reflections to an R factor of 27.4% and an R-free value of 29.7%. All missing residues, as well as side chains, were built into the model. Two well-defined electron densities that matched PL or a derivative of PL, each at the active sites of the monomers, were observed and built into the model. Densities with peaks above 2.5σ and 1.0σ in the Fo-Fc and 2Fo-Fc maps, respectively, and within acceptable hydrogen-bonding distances were added to the model as water molecules. Repeated cycles of refinements, including composite omit map and model building with the graphics programs TOM (2) and XtalView (11), resulted in a model that included residues 1 to 287 for each monomer, 2 PL molecules, 1 sulfate molecule, and 385 water molecules, with an R factor and an R-free value of 17.7 and 24.6%, respectively. The refinement statistics are summarized in Table 1. The structural coordinates have been deposited in the PDB under code 1TD2.

RESULTS AND DISCUSSION

Overall structure of PdxY. The crystal structure of PdxY is reported at a resolution of 2.22 Å and was solved by a molecular replacement method (13), with sheep PL kinase (PDB

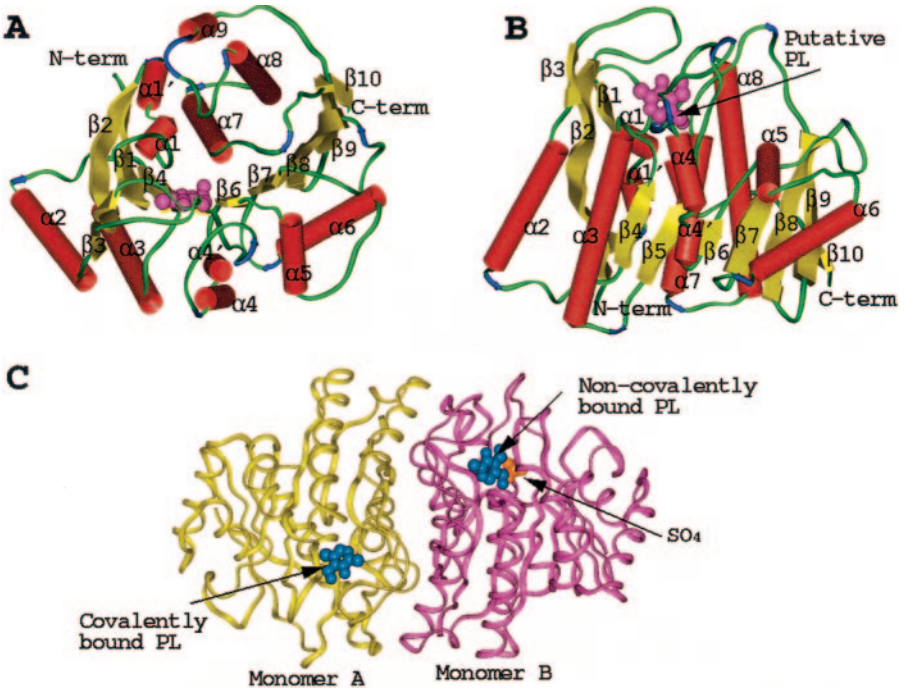

FIG. 1. Ribbon diagram of PdxY fold. (A and B) Orthogonal views of the monomeric structure. The right view is rotated 90° from the left view. α-Helices and β-strands are shown in yellow and red, respectively. The secondary structures are labeled. The bound putative PL molecule is shown with magenta spheres. (C) Dimeric structure showing the covalently and noncovalently bound putative PL (blue spheres) at monomer A (yellow) and monomer B (magenta), respectively. The bound sulfate at monomer B is also shown with orange spheres. The figures were generated with Insight II (Molecular Simulations, Inc., San Diego, Calif.) and were labeled with Showcase.

code 1LHR) as the search model. PdxY and sheep PL kinase share 45% sequence similarity and 31% sequence identity, and as expected the two proteins have a very similar fold. Refinements and other crystallographic parameters are listed in Table 1. The electron density map (2Fo-Fc) at 1.0σ shows a well-defined density for almost the entire polypeptide of main-chain atoms, except for weak densities for residues 117 to 122 of monomer A, which are part of a loop region at the active site of the enzyme. The R-merge value was significantly high, which was probably due to weak diffraction as well as a gradual deterioration of the crystal during data collection. The crystallographic asymmetric unit contains a homodimer, with the two monomers related by a noncrystallographic twofold axis (Fig. 1). Each monomer consists of 11 α-helices, 3 3_{10} helices, and 10 β-strands. The strands form a central core flanked by the helical structures, similar to the ribokinase superfamily fold (3, 4). Both subunits can be superimposed, with a root mean square deviation (RMSD) of 0.55 Å for all 287 pairs of Cα atoms, with monomer A on monomer B and vice versa. If the matching criterion of 1.0 Å is imposed, the 276 matched atom pairs deviate by an RMSD of 0.28 Å, with the exception of residues 116 to 121 and 190 to 191. The former deviation is of the most significance and corresponds to a conformational change of the loop from residues 116 to 127 over the active site (see below). It appears that crystal contact interactions (mostly hydrophobic and water-mediated hydrogen bonds) limit monomer B's loop to a stable closed position. Conversely, the lack of crystal packing interactions resulted in high B factors for monomer A's loop and allowed it to rotate slightly, although still in a closed position. As will be discussed later, this loop is closely associated with the bound putative PL(P) at the active site. The buried surface area calculated with the CNS program suite (1) was 3,410 $Å^2$ for the dimer.

Structural comparison of PdxY and 1VI9. The crystal structure of the *E. coli pdxY* gene product has also been determined by Structural Genomix and was deposited in the PDB as 1VI9. In the subsequent text, this structure will be referred to as 1VI9 to differentiate it from the PdxY structure described in this paper. Since the structure of PdxY was solved prior to the deposition of 1VI9, we used the sheep PL kinase structure for molecular replacement analysis. Both PdxY and 1VI9 crystallized in a monoclinic cell (space group $P2_1$). However, the cell volume of 1VI9 was twice that of PdxY, resulting in two independent dimer molecules (dimers AB and CD) in 1VI9 compared to a single dimer (dimer AB) in PdxY in the asymmetric unit. The two dimers of 1VI9 superimposed each other (using all 288 pairs of Cα atoms), with an RMSD of 0.28Å. In contrast, PdxY superimposed both dimers AB and CD, with an RMSD of ~1.2 Å. The RMSD ranged from 0.4 to 1.4 Å when monomers A and B of PdxY were independently compared with the four monomers of 1VI9 (using all Cα atoms), with the most significant deviation located at the loop regions from residues 116 to 127 and 48 to 54 over the active site. Interestingly, the larger deviations occurred between the monomers of PdxY and between monomers A and C of 1VI9, in which the two loops were significantly rotated away to open up the active site in 1VI9. This was in contrast to the loops in monomers B and D of 1VI9 as well as those in PdxY, which were found in closed positions.

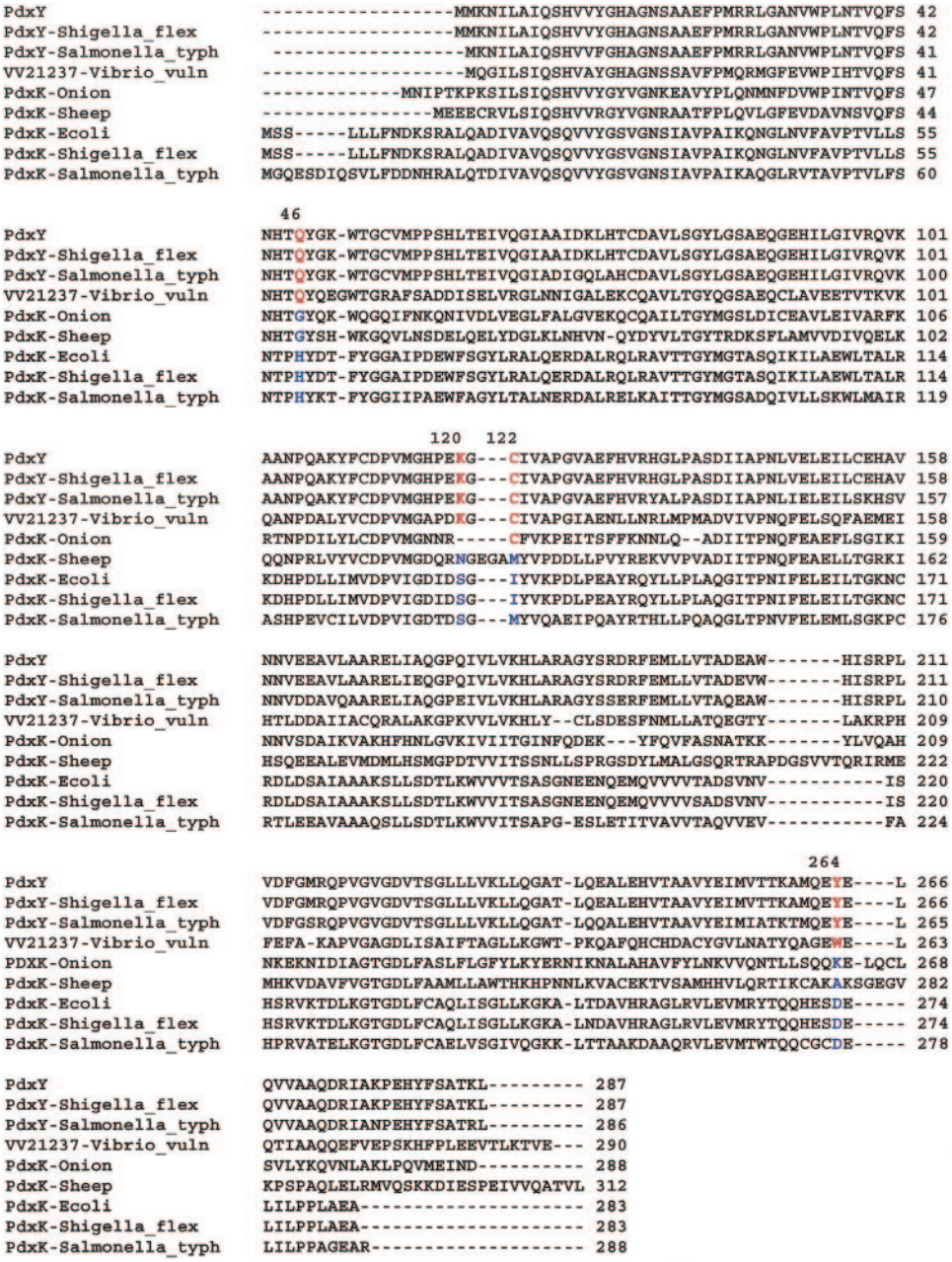

FIG. 2. Multiple sequence alignment of 9 of the 34 putative PL kinases from a sequence homology search with the FASTA program (15). The proteins used for the sequence alignment were as follows (database numbers are shown in parentheses): PdxY, *E. coli* PdxY (P77150); PdxY-Shigella_flex, *Shigella flexneri* PdxY (Q83KY1); PdxY-Salmonella_typh, *Salmonella enterica* serovar Typhimurium PdxY (Q8ZPM8); VV21237-Vibrio_vuln, *V. vulnificus* VV21237 protein (Q8D4Q2); PdxK-Onion, onion PdxK (BAD04519); PdxK-Sheep, sheep PdxK (P82197); PdxK-Ecoli, *E. coli* PdxK (P40191); PdxK-Shigella_flex, *Shigella flexneri* PdxK (Q7UC31); PdxK-Salmonella_typh, *Salmonella enterica* serovar Typhimurium PdxK (P40192). The sequences PdxY-Shigella_flex, PdxY-Salmonella_typh, and VV21237-Vibrio_vuln have high sequence identities with PdxY (46 to 99%), while the last five sequences have low sequence identities with PdxY (28 to 41%). Shown in red are the conserved active site residues, Gln46, Lys120, Cys122, and Tyr264, in the sequences that have high sequence identities with PdxY. The corresponding nonconserved residues in the other PL kinase members are shown in blue. The figure was generated with ClustalW in the FASTA program (15).

Sequence comparison of PdxY and other PL kinases. A sequence homology search with the FASTA program (15) found 34 PL kinases (including PdxY) with sequence identities to PdxY, ranging from 29 to 99%. Interestingly, the 13 sequences with the highest sequence identities to PdxY (46 to 99%) have four unique residues (Gln46, Lys120, Cys122, and Tyr264) that are conserved at the active site. The only exceptions are sequences for *Vibrio vulnificus* and *Vibrio parahaemolyticus*, which have Tyr264 replaced with Trp. These two gene products have the lowest sequence identity (46%) to PdxY among the 13 most related sequences. In the other 21 PL kinase sequences, which have <45% sequence identities to PdxY (29 to 41%), including those for the sheep (30%) and human (29%) enzymes and *E. coli* PL kinase 1 (30%), none of

the above four active site residues is conserved, with the exception of an onion enzyme (35%) which has a conserved Cys122 residue. In these last sequences, Gln46 was replaced mainly with His, Gly, or Phe, Lys120 was replaced mainly with Gly, Ser, or Thr, Cys122 was replaced with hydrophobic residues, and Tyr264 was replaced mainly with Glu, Asp, or Leu. Due to the large number of sequences, we only show representatives from the two classes of sequences mentioned above (Fig. 2). Interestingly, like *E. coli*, which has two isoform genes (*pdxY* and *pdxK*), *Shigella flexneri* and *Salmonella enterica* serovar Typhimurium also have two such genes, with one isoform gene (*pdxY*) encoding a protein with >90% sequence identity with PdxY and the other gene (*pdxK*) encoding a protein with <30% sequence identity with PdxY but 66% or more identity with PL kinase 1 (66 and 98% for *Shigella flexneri* and *Salmonella enterica* serovar Typhimurium, respectively). Of all 34 PL kinases, the crystal structures of only PdxY and sheep PL kinase are known.

Structural comparison of PdxY and members of the ribokinase superfamily. A database search with Fugue (17) found six structural homologues, including sheep PL kinase (PDB codes 1LHR and 1LHP), with a *z* score of 36%; 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMPP) kinase from *Salmonella enterica* serovar Typhimurium (PDB code 1JXH), with a *z* score of 32%; 4-methyl-5-β-hydroxyethylthiazole (THZ) kinase from *Bacillus subtilis* (PDB code 1C3Q), with a *z* score of 20%; a hypothetical protein in the Sigy-Cydd intergenic region with an unknown function (PDB code 1KYH), with a *z* score of 19; human adenosine (HAD) kinase (PDB code 1BX4), with a *z* score of 7%; and ribokinase (eRK) from *E. coli* (PDB code 1RKD), with a *z* score of 7%. All six proteins are members of the ribokinase superfamily, with the same central core structure of β-sheets surrounded by α-helices. With the exception of sheep PL kinase, none of these proteins uses PL/PM/PN as a substrate. The amino acid identities between PdxY and the non-PL kinases are very low (<30%). Like the 21 PL kinases that have <45% sequence identity to PdxY, none of the four residues (Gln46, Lys120, Cys122, and Tyr264) found at the active site of PdxY are conserved in these non-PL kinases. The only exception is HAD kinase, which appears to have a conserved Cys122, but this Cys is far removed from the active site.

A monomer of PdxY can be superimposed on sheep PL kinase with an RMSD of 1.4 Å for 269 pairs of Cα atoms. For comparison, the RMSD is 2.1 Å for HMPP kinase (228 pairs of Cα atoms), 2.3 Å for eRK (227 pairs of Cα atoms), 2.7 Å for THZ kinase (216 Cα atoms), and 3.9 Å for HAD kinase (175 Cα atoms). PdxY is a dimer in the crystal and is comparable to sheep PL kinase, HMPP, and eRK, which also function as homodimers. Comparisons of the dimers showed that PdxY has a quaternary structure that is closest to those of the sheep PL (RMSD of 1.7 Å) and HMPP (RMSD of 2.3 Å) kinases and significantly different from that of eRK (RMSD of 4.2 Å). In contrast, HAD kinase is a monomer and THZ kinase is a trimer. Comprehensive analyses of the ribokinase superfamily of proteins have been reported by Cheng et al. (4) and Campobasso et al. (3).

Active site. There are two putative active sites in the dimeric structure of PdxY, and each is exclusively formed by residues from one monomer. The active site starts as a shallow groove and stretches to the backside of the dimer interface. The putative PL(P) site is located at the deeper end of the active site, and it appears certain that if ATP binds it does so at the shallow groove, close to the mouth of the active site cavity (see below). The active site geometry is conserved in the ribokinase superfamily, and that of PdxY is especially close to those of the sheep PL and HMPP kinases. Like PdxY, each active site of the sheep PL, HMPP, and HAD kinases is located entirely within a monomer. In contrast, the active sites of THZ kinase and eRK are formed at the interface of two monomers.

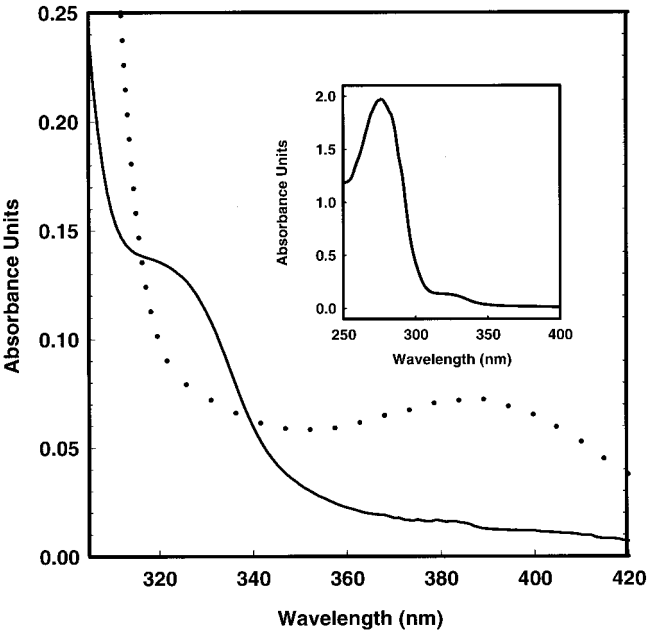

FIG. 3. Spectrum of purified PdxY (2 mg/ml, in 20 mM potassium phosphate, pH 7.3). The solid line shows the spectrum for PdxY between 300 and 420 nm. The dotted line is the spectrum after adding NaOH to 0.1 M, showing the shift of most of the 325-nm peak to one that absorbs at 388 nm, which is characteristic of PLP. Inset, spectrum of PdxY showing both the 278- and 325-nm peaks.

PL binding site. The PL binding cavity is formed by the C terminus of strand β1, the loop between strand β1 and helix α1, the C terminus of strand β2, the loop between strands β2 and β3, the loop between strand β4 and helix α3, the loop between strand β5 and helix α4, the loop between strand β9 and helix α7, and the N terminus of helix α7. Unexpectedly, two fully occupied ligands were observed bound at the active site of each monomer in PdxY, even though the enzyme was purified and crystallized in the absence of any specific ligands. As discussed below, both the spectral studies and the release of PL and PLP upon denaturation are consistent with the density being attributed to PL and/or a mixture of PL and PLP (Fig. 3 and Table 2). The electron density fits well to a PL moiety being bound, possibly as a thiohemiacetal, at one active site (monomer A) and a PL and a sulfate molecule being bound at the second active site (monomer B). The N-terminal end of helix α7 and its positive dipole are directed toward O-3′ of PL and the observed sulfate molecule. The structure was refined to a PL thiohemiacetal (monomer A) and PL (monomer B). Figures 4A and C show different Fourier electron density maps (using phases before PL was built into the model), and Fig. 4B and D show omit maps, all of which were calculated around the binding sites of the PL in both monomers, with specific interactions between PL and monomer A shown in Fig. 5A and B.

TABLE 2. Stoichiometry of PL and PLP bound to PdxY

Protein	Amt of PLP (pmol) or ratio of PLP to PdxY (mean ± SD)		
	−MgATP	+MgATP	+MgATP and ePL kinase
PdxY denatured with heat[a]			
PLP	62.7	84.1	128
PLP/PdxY	0.24 ± 0.01	0.31 ± 0.02	0.48 ± 0.01
PdxY denatured with acid and heat[b]			
PLP	75.7 ± 0.01	83.3	131
PLP/PdxY	0.28 ± 0.01	0.31 ± 0.02	0.49 ± 0.00

[a] A total of 269 pmol of PdxY.
[b] Denatured at pH 3.3 and 100°C.

There is an apparent covalent bond between C-4′ of the putative PL and the sulfhydryl group (SG) of Cys122 in monomer A, with a bond distance of 2.0 Å and a strong overlapping density (Fig. 4A and B). The electron density map also shows the bond between C-4′ of PL and the SG of Cys122 to be outside of the plane of the pyridoxal ring, suggesting that C-4′ is tetrahedral. The density also fits well with an out-of-plane -OH group on C-4′, indicating that PL may be bound as a thiohemiacetal. In contrast, at the active site of monomer B, the bound PL does not seem to make any covalent interaction with Cys122, as indicated by the relatively long distance of 3.6 Å between C-4′ and the SG as well as a lack of overlapping density between the two atoms (Fig. 4C and D). Also, the omit map shows the oxygen on C-4′ to be in the plane with the PL ring. The sulfate that binds at the active site of monomer B lies close to the -OH of C-5′ of PL. We did not observe any overlapping density between the sulfate molecule and PL, but one of the sulfate oxygen atoms came as close as ~2.9 Å to C-5′. Also, the C-5′ -OH of the PL in monomer B was found to occupy two positions, and in one position (not refined), this -OH came as close as ~2.8 Å to the sulfate sulfur atom. In addition to the hydrogen bond interaction with the C-5′ -OH of PL, the sulfate also makes hydrogen bond interactions with the defined anion-hole residues of Gly221, Val222, Gly223, and Asp224. These residues, with the exception of Val222, are highly conserved in the PL kinases and other members of the ribokinase superfamily.

In contrast, at monomer A, where a covalent interaction is observed between PL and Cys122, the sulfate position is occupied by a water molecule. We assumed the presence of a water molecule because the density was very weak and was not tetrahedrally shaped. This water molecule also interacts with the C-5′ -OH of PL. While PL appears to be covalently bound to subunit A, we hypothesize that the bound ligand at monomer B is a mixture of PL and PLP that are noncovalently bound to the protein. The stoichiometry for PLP (0.6) (as discussed below) comes close to the approximately 72% crystallographic occupancy observed for the sulfate ligand in the structure.

Interestingly, there is a β-mercaptoethanol molecule covalently bound to Cys122 at the active sites of monomers B and D (one each from the two independent dimers) of 1VI9, while the other two monomers (A and C) lack any bound ligand. The PL(P) and β-mercaptoethanol molecules in PdxY and 1VI9, respectively, can superimpose each other. Sulfate ions are also

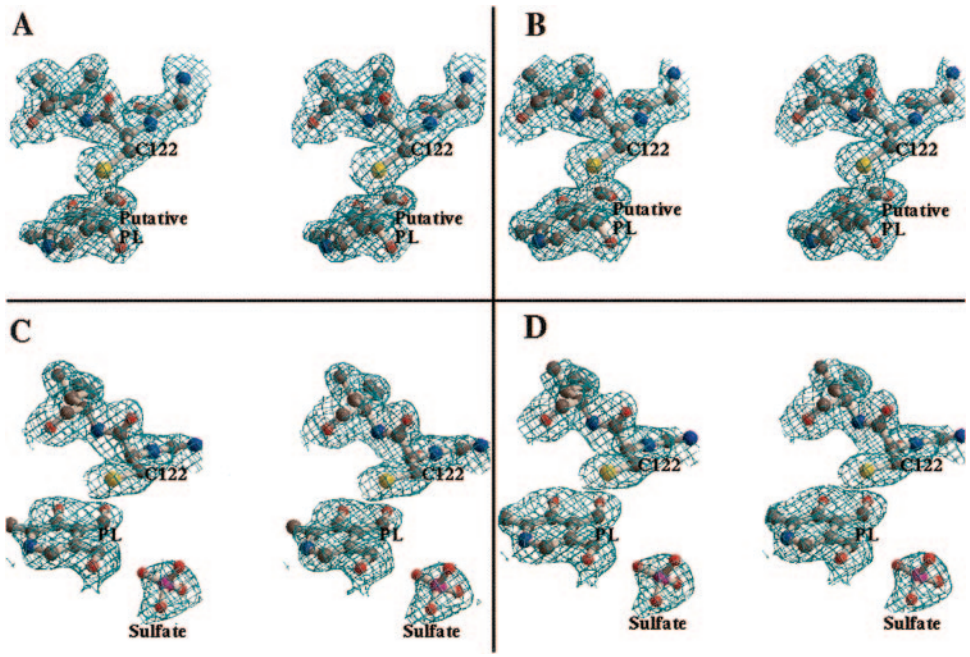

FIG. 4. Stereoviews of electron density map of bound putative PL at the active site of PdxY. (A) 2Fo-Fc map of the covalently bound derivative of PL at monomer A, which was calculated before PL was built into the model. (B) 2Fo-Fc omit map of the covalently bound derivative of PL at monomer A, which was calculated by omitting PL during simulated annealing. (C) 2Fo-Fc map of the noncovalently bound PL at monomer B, which was calculated before PL was built into the model. (D) 2Fo-Fc omit map of the noncovalently bound PL at monomer B, which was calculated by omitting PL during simulated annealing. The maps were contoured at the 1.0σ level and then superimposed with the final refined model. The figures were drawn with Bobscript (6) and Raster3D (12) software and were labeled with Showcase (Silicon Graphics, Inc.).

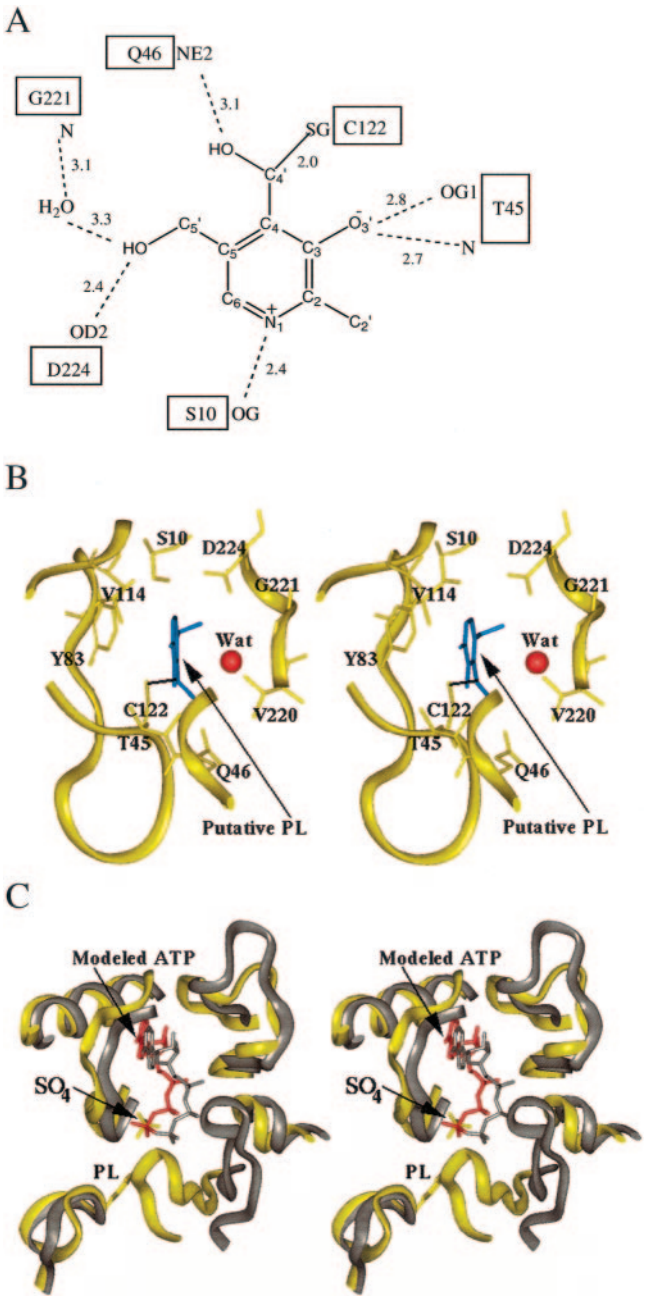

FIG. 5. Active site of PdxY. (A) Schematic diagram showing the hydrogen bond interactions between the putative PL and the active site residues of the monomer. Hydrogen bonds are shown as dashed lines, and the observed covalent bond between C-4′ of PL and the SG of Cys122 is shown as a solid line. (B) Stereoview of the putative PL binding site at monomer A. The protein is shown with yellow ribbon figures. The PL and the active site residues are shown as cyan and yellow sticks, respectively. Water molecules are shown as red spheres. The covalent interaction between C-4′ of PL and the SG of Cys122 is shown with a line. For clarity, not all of the active site amino acid residues are shown. (C) Superposition of the active sites of PdxY (yellow) and sheep kinase (gray). ATP, PL, and sulfate molecules are shown with stick models. The ATP in red is the theoretical model for PdxY. The figures were generated with Insight II and labeled with Showcase.

found at the active sites of all four monomers of 1VI9, located in the same position as the sulfate in PdxY.

The PdxY protein was shown to have only about 0.1% PL kinase activity compared to the PL kinase 1 protein that is also present in *E. coli* (5, 20). Once the crystallized protein was found to have a mixture of tightly bound PL and PLP at one site and a putative derivative of PL covalently attached at the second site, we were interested in determining how the active site residues that interact with PL in PdxY compare to the residues in other PL kinases. N-1, O-3′, the oxygen on C-4′, and the -OH on C-5′ of PL have hydrogen bond interactions with Ser10, Thr45, Gln46, and Asp224, respectively (Fig. 5A). There is a π-π stack interaction between the PL ring and Tyr83 that may help to orient the molecule for proper interaction and catalysis. Other hydrophobic contacts include Phe41, Val114, and Val220. Similar interactions were observed at the monomer B active site. The residues Ser10, Val114, Asp224, Phe41, and Tyr83 are almost totally conserved among all PL kinases, and it has been suggested that Ser10, Tyr83, and Asp224 determine substrate specificity (9). Cys122, Gln46, and Val220 are totally conserved in those PL kinases that have high sequence identities with PdxY (>45%), and remarkably they always occur together in these kinases (Fig. 2). In contrast, in the putative PL kinases with low sequence identities with PdxY (Fig. 2), as well as the members of the ribokinase superfamily with known crystal structures, both Cys122 and Gln46 are absent from the active site, thus precluding these enzymes from making the covalent interactions with PL that were observed for PdxY. Interestingly, Gln46 has only one hydrogen bond interaction with the C-4′ -OH of PL in monomer A of PdxY, but in monomer B (where there is no apparent covalent interaction between PL and Cys122) Gln46 is disordered in two positions, with one position having an interaction with the oxygen of C-4′ and the other having hydrogen bond interactions with Lys120 and Tyr264. As pointed out above, with the exception of onion PL kinase (which also has a Cys residue at position 122), none of the four conserved residues (Gln46, Lys120, Cys122, and Tyr264) is present and there are not even any similarities to these residues in the other kinases that have sequence identities of <45% with PdxY. We speculate that Gln46 and Cys122 are essential for the function of the PdxY family of proteins. It is also possible that Lys120 and Tyr264, even though they make no direct contact with the substrate, play an indirect role in the function of the enzyme. We should point out that none of these putative PL kinases that are related to PdxY have been isolated and shown to be catalytically active. They are listed as PL/PN/PM kinases in the protein sequence databases because of their sequence similarities.

Flap covering the active site. A characteristic feature of ribokinase structures is a flap or lid that covers the substrate binding site, which has been used to propose an evolutionary pathway for the enzymes in the ribokinase superfamily (4). In PdxY (or 1VI9), as well as in HMPP and sheep PL kinases, this flap is formed by the loops at residues 116 to 127 (between strand β5 and helix α4) and 42 to 51 (between strands β2 and β3) from the same monomer (residues correspond to that of PdxY). In PdxY, the loops at residues 116 to 127 and 42 to 51 that form the active site flap are in closed positions in both monomers. In contrast, in 1VI9, the loops in two of the monomers (monomers A and C) are in open conformations, while

they are in closed conformations in the other two monomers. Interestingly, the active sites with closed loop conformations have bound ligands [PL(P) in PdxY and β-mercaptoethanol in 1VI9] and vice versa. It therefore appears that the conformational change of the loop is related to binding of the substrate, which triggers the closure of the active site.

In HMPP, the corresponding loop residues 116 to 127 also show some flexibility, and like that for PdxY (1VI9), the crystal structure shows the loop opening and closing upon substrate binding (4). The flexibility of this loop is even more apparent for sheep PL kinase, which has an additional three-residue insertion in the loop, making it significantly longer than those of PdxY and HMPP kinases. In contrast to the case for both PdxY and HMPP, in the sheep PL kinase structure without ATP and a bound substrate (PDB code 1LHP), the loop occupies the putative substrate position, while in one of the monomers of a sheep PL kinase-ATP complex structure without a bound substrate (PDB code 1LHR), the loop has swung out of the active site to assume an open position that clearly opens the active site to solvent and access to PL. Also, in the sheep PL kinase, Tyr127 and Asp118 (from the loop that occupies the putative substrate position) form hydrogen bonds with the γ- and β-phosphates of ATP (9). Tyr127 and Asp118 are almost completely conserved in the 21 PL kinases (including *E. coli* PL kinase 1, the isoform of PdxY) that have <45% sequence identity with PdxY. On the other hand, these two residues are replaced with Ile and His in PdxY and the other PL kinases that have >45% sequence identity with PdxY. The ability of the loop to assume closed and open positions in sheep PL kinase, as well as the presence of the conserved Tyr127 and Asp118, led the authors to propose that the function of the loop in these kinases is to allow access of PL to the substrate binding site independent of a bound ATP as well as to interact with ATP to prevent premature ATP hydrolysis, consistent with a random sequential kinetic mechanism (9, 10). In contrast, for HMPP, which has nonconserved Tyr127 and Asp118 residues, this loop is believed to act as a flap to cover the bound non-ATP substrate (4), and unlike the sheep PL kinase, it is suggested that the substrate can only bind in the absence of ATP. Hence, this kinase shows an ordered addition of substrate. The absence of Tyr127 and Asp118 in the loop of PdxY, as well as the structural similarities of the loop conformation with that of HMPP, suggests that the flap in PdxY as well as those in the other PL kinases with high sequence identities with PdxY may function similarly to that of HMPP.

Putative ATP binding site. The crystal structure presented here has no bound ATP, but a comparison of the active site structures of PdxY and the other members of the ribokinase superfamily that have bound ATP show identical ATP binding geometries. Figure 5C shows a modeled ATP at the active site of PdxY based on the sheep PL kinase structure. The modeled ATP is located in a binding pocket formed by the loop between strand β5 and helix α4, the loop between strand β6 and helix α5, the loop between strands β7 and β8, the N terminus of strand β8, the N terminus of helix α7, the loop between strand β9 and helix α7, and the middle segment of helix α8. The predicted position of the γ-phosphate overlaps the observed sulfate ion positions in both PdxY and 1VI9. N-1, N-6, and N-7 of the adenosine ring have possible hydrogen bond interactions with the residues Gln218 and Val212, while residue Phe194 can be rearranged from its current position to form a π-π stacking interaction with the adenosine ring. The sugar moiety has mainly hydrophobic interactions with Phe194, Val222, and Met196 and a hydrogen bond interaction with Thr226. The residues His183 and Lys182 can also be rearranged from their current positions to interact with the α- and β-phosphate groups. Lastly, the Asp112, Asn146, His117, Gly221, Val222, Gly223, Asp224, and Glu149 residues can have direct hydrogen bond or water-mediated hydrogen bond interactions or can be rearranged to form hydrogen bond interactions with the γ-phosphate.

All the above residues that potentially interact with ATP are almost totally conserved in the 13 putative PL kinases that have all four of the active site residues (Cys122, Lys120, Tyr264, and Gln46). In contrast, only Asp112, Asn146, Glu149, Gly221, Gly223, and Asp224 are highly conserved in all 34 PL kinases, but these residues are mostly absent from the other members of the ribokinase superfamily. In addition to the sequence motif GVGA (residues 221 to 224) that forms the anion hole and interacts with the γ-phosphate of the modeled ATP, there is another highly conserved sequence motif, DPVMG (residues 112 to 116), that runs almost parallel to the GVGA motif in the pyridoxal kinases. The γ-phosphate is positioned between these two sequence motifs but is closer to the GVGA motif. As noted above, all four residues from the GVGA motif and Asp112 from the DPVMG motif form hydrogen bond and water-mediated hydrogen bond interactions, respectively, with the γ-phosphate. We should point out that the modeled ATP γ-phosphate overlaps with the bound sulfate in both PdxY and 1VI9.

Evidence for PL and PLP bound to the purified protein. The crystal structure of PdxY suggests that there is bound PL(P) at the two active sites. To test this hypothesis, we performed the following experiments. The PdxY protein was purified to homogeneity as previously described (5). The spectrum showed an absorption peak at 325 nm in addition to the normal absorption peak at 278 nm (Fig. 3). Denaturation of the protein by the addition of NaOH resulted in a shift of the 325-nm peak to 388 nm, which is the absorption maximum of PLP (Fig. 3). This suggested that the 325-nm-absorbing compound was released as PLP when the protein was denatured. To confirm that PLP is tightly bound to PdxY, we denatured the protein by boiling it either at a neutral pH or at pH 3.3 (to release any adducts of PLP) and analyzed the supernatant for released PLP. The analysis was achieved by converting inactive aposerine hydroxymethyltransferase (a PLP-requiring enzyme) to the active holoenzyme (7). This was also done with aliquots of PdxY that had been incubated with MgATP to convert any PL bound to the enzyme to PLP prior to denaturation. The results showed that about 0.25 equivalent of PLP was released per monomer of PdxY by heat denaturation in the absence of MgATP (Table 2) and about 0.3 equivalent of PLP was released per monomer of PdxY that had been incubated with MgATP prior to denaturation (Table 2). To determine if PL had been released by the denaturation of PdxY, we incubated the supernatant with MgATP and PL kinase 1 before adding aposerine hydroxymethyltransferase (Table 2). Under these conditions, we found about 0.5 equivalent of PLP per monomer of PdxY. The results suggest that upon denaturation of the enzyme, about 0.3 equivalent of PLP and another 0.2 equivalent of PL are released per monomer. It also appears that

incubating PdxY with MgATP prior to denaturation did not convert the bound PL to PLP. We also tested for pyridoxine and pyridoxamine, but no increase in the amount of PLP liberated was observed, suggesting that these two vitamin forms are not bound to PdxY. These results show that PL and PLP are bound so tightly to PdxY that they are not released during purification, which included several ammonium sulfate precipitations, extensive dialyses, and ion-exchange chromatography.

Because only 1 equivalent of combined PLP and PL was released per dimer from denatured PdxY, we tested for a covalently bound derivative of PL remaining with the denatured protein. PdxY was denatured in either sodium dodecyl sulfate, 6 M urea, 6 M guanidine HCl, or 0.1 M NaOH and then chromatographed down a P6-DG molecular sieve column. Under each condition, some of the 325-nm-absorbing material remained with the protein, suggesting a covalent attachment of at least some of the chromophore.

Function of PdxY. PdxY clearly belongs to the ribokinase superfamily. However, the conserved residues at the ATP binding site, the unique active site flap, and the exclusive presence of all four active site residues, Gln46, Lys120, Cys122, and Tyr264, in some of the related kinases, as well as the apparent covalent interaction between PL and Cys122, suggest that these kinases may function differently from other kinases that lack these four residues. We propose that the proteins that have high levels of sequence similarity with PdxY represent a new structural family in the ribokinase superfamily of proteins. However, the function of PdxY and its related proteins (most of them also from the *pdxY* gene) remains unknown. PdxY was discovered as a possible PL kinase in *E. coli* because it supported growth on PL when *pdxK* and the biosynthetic pathway genes were inactivated. At this point, we have been able to find homologues of PdxY only in other prokaryotic organisms, but we cannot rule out that members of this group also exist in eukaryotic cells.

Previous studies have suggested that because of the very low catalytic activity of PdxY with PL as a substrate, this protein is a kinase for another metabolic pathway but has enough activity with PL to support *E. coli* growth (5, 20). The structure of PdxY casts doubt on this interpretation because the protein that we isolated from *E. coli* has both active sites filled by very tightly bound ligands. At least half of these ligands are a mixture of PL and PLP (Table 2), and the other half are most likely a covalently attached form of PL. If the sites are filled with tightly bound ligands, then it is unlikely that the protein functions in another metabolic pathway. A previous study showed that *E. coli* extracts contained PL that was tightly bound to a protein and noted that no protein had previously been shown to tightly bind PL (7). Now we can say that PdxY can account for this previous observation of PL being bound to a protein in *E. coli* extracts.

Why do PLP and PL stick so tightly at the noncovalent site? From the structure of the sheep PL kinase, it was proposed that a flap covers the PL subsite during catalytic transfer of the γ-phosphate of ATP to the C-5′ -OH of PL. This would protect the transfer from solvent and the premature hydrolysis of ATP. In PdxY, this flap covers the bound PL(P) and sequesters it from the solvent, even in the absence of bound ATP. Apparently, in PdxY this property of a closed structure accounts for the tight binding of the ligand. Further studies are required to determine how the flap can be moved in monomer B to a position for the release of bound PL(P) and to understand the nature of the covalently attached group in monomer A.

ACKNOWLEDGMENTS

This work was supported by NIH grant DK56648. Data for this study were measured at beamline X26C of the National Synchrotron Light Source. Financial support came principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the U.S. Department of Energy and from the National Center for Research Resources of the National Institutes of Health.

We thank Annie Heroux for her help in obtaining the data set.

REFERENCES

1. **Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren.** 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D **54:**905–921.
2. **Cambillau, C., and E. Horjales.** 1987. TOM: a Frodo subpackage for protein-ligand fitting with interactive energy minimization. J. Mol. Graph. **5:** 174–177.
3. **Campobasso, N., I. I. Mathews, T. P. Begley, and S. E. Ealick.** 2000. Crystal structure of 4-methyl-5-beta-hydroxyethylthiazole kinase from Bacillus subtilis at 1.5 Å resolution. Biochemistry **39:**7868–7877.
4. **Cheng, G., E. M. Bennett, T. P. Begley, and S. E. Ealick.** 2002. Crystal structure of 4-amino-5-hydroymethyl-2-methylpyrimidine phosphate kinase from *Salmonella typhimurium* at 2.3 Å resolution. Structure **10:**225–235.
5. **Di Salvo, M. L., S. Hunt, and V. Schirch.** 2004. Expression, purification and kinetic constants for human and Escherichia coli pyridoxal kinases. Protein Expr. Purif. **36:**300–306.
6. **Esnouf, R. M.** 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. J. Mol. Graph. **15:**132–134.
7. **Fu, T.-F., M. di Salvo, and V. Schirch.** 2001. Distribution of B6 vitamers in *Escherichia coli* as determined by enzymatic assay. Anal. Biochem. **298:** 314–321.
8. **Laine-Cessac, P., A. Cailleux, and P. Allain.** 1997. Mechanisms of the inhibition of human erythrocyte pyridoxal kinase by drugs. Biochem. Pharmacol. **54:**863–870.
9. **Li, M.-H., F. Kwok, W.-R. Chang, S.-Q. Liu, S. C. L. Lo, J.-P. Zhang, T. Jiang, and D.-C. Liang.** 2004. Conformational changes in the reaction of pyridoxal kinase. J. Biol. Chem. **279:**17459–17465.
10. **Li, M. H., F. Kwok, W. R. Chang, C. K. Lau, J. P. Zhang, S. O. Liu, Y. C. Leung, T. Jiang, and D. C. Liang.** 2002. Crystal structure of brain pyridoxal kinase, a novel member of the ribokinase superfamily. J. Biol. Chem. **277:** 46385–46390.
11. **McRee, D. E.** 1999. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. **125:**156–165.
12. **Merrit, E. A., and M. E. P. Murphy.** 1994. Raster3D version 2.0: a program for photorealistic molecular graphics. Acta Crystallogr. D **50:**869–873.
13. **Navaza, J.** 1994. AMoRe: an automated package for molecular replacement. Acta Crystallogr. D **50:**157–163.
14. **Otwinowski, Z., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276:**307–326.
15. **Pearson, W. R.** 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. **183:**63–98.
16. **Perrakis, A., M. Harkiolaki, K. S. Wilson, and V. S. Lamzin.** 2001. ARP/wARP and molecular replacement. Acta Crystallogr. D **57:**1445–1450.
17. **Shi, J., T. L. Blundell, and K. Mizuguchi.** 2001. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J. Mol. Biol. **310:**243–257.
18. **Sivaraman, J., Y. Li, J. Banks, D. E. Cane, A. Matte, and M. Cygler.** 2003. Crystal structure of *Escherichia coli* PdxA, and an enzyme involved in the pyridoxal phosphate biosynthesis pathway. J. Biol. Chem. **278:**43682–43690.
19. **Van Asselt, E. J., A. Perrakis, K. H. Kalk, V. S. Lamzin, and B. W. Dijkstra.** 1998. Accelerated X-ray structure elucidation of a 36 kDa muramidase/transglycosylase using wARP. Acta Crystallogr. D **54:**58–73.
20. **Yang, Y., H. C. Tsui, T. K. Man, and M. E. Winkler.** 1998. Identification and function of the *pdxY* gene, which encodes a novel pyridoxal kinase involved in the salvage pathway of pyridoxal 5′-phosphate biosynthesis in *Escherichia coli* K-12. J. Bacteriol. **180:**1814–1821.
21. **Yang, Y., G. Zhao, and M. E. Winkler.** 1996. Identification of the *pdxK* gene that encodes pyridoxine (vitamin B_6) kinase in *Escherichia coli* K-12. FEMS Microbiol. Lett. **141:**89–95.