Molecular Structure of *Escherichia coli* PurT-Encoded Glycinamide Ribonucleotide Transformylase[†,‡]

James B. Thoden,[§] Steven Firestine,[∥] Andrew Nixon,[⊥] Stephen J. Benkovic,[∥] and Hazel M. Holden*[,§]

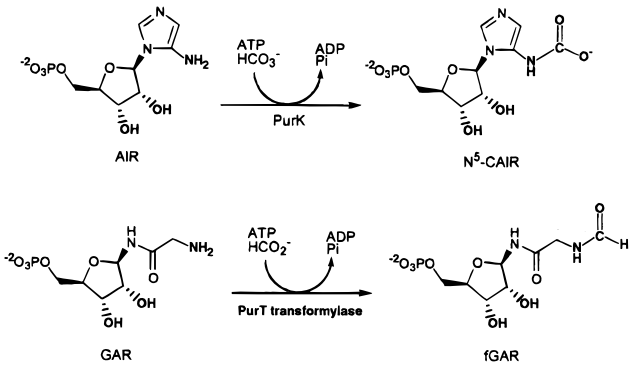
Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53705, Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, and Dyax Corporation, One Kendall Square, Building 600, 5th Floor, Cambridge, Massachusetts 02139

Received April 24, 2000; Revised Manuscript Received May 24, 2000

ABSTRACT: In *Escherichia coli*, the PurT-encoded glycinamide ribonucleotide transformylase, or PurT transformylase, catalyzes an alternative formylation of glycinamide ribonucleotide (GAR) in the de novo pathway for purine biosynthesis. On the basis of amino acid sequence analyses, it is known that the PurT transformylase belongs to the ATP-grasp superfamily of proteins. The common theme among members of this superfamily is a catalytic reaction mechanism that requires ATP and proceeds through an acyl phosphate intermediate. All of the enzymes belonging to the ATP-grasp superfamily are composed of three structural motifs, termed the A-, B-, and C-domains, and in each case, the ATP is wedged between the B- and C-domains. Here we describe two high-resolution X-ray crystallographic structures of PurT transformylase from *E. coli*: one form complexed with the nonhydrolyzable ATP analogue AMPPNP and the second with bound AMPPNP and GAR. The latter structure is of special significance because it represents the first ternary complex to be determined for a member of the ATP-grasp superfamily involved in purine biosynthesis and as such provides new information about the active site region involved in ribonucleotide binding. Specifically in PurT transformylase, the GAR substrate is anchored to the protein via Glu 82, Asp 286, Lys 355, Arg 362, and Arg 363. Key amino acid side chains involved in binding the AMPPNP to the enzyme include Arg 114, Lys 155, Glu 195, Glu 203, and Glu 267. Strikingly, the amino group of GAR that is formylated during the reaction lies at 2.8 Å from one of the γ-phosphoryl oxygens of the AMPPNP.

In higher organisms, the de novo purine biosynthetic pathway consists of 10 enzymatic steps, while in bacteria, plants, fungi, and yeast, 11 steps are required (*1*). The additional enzyme in lower organisms is N^5-carboxyaminoimidazole ribonucleotide synthetase, hereafter called PurK, which catalyzes the ATP-dependent carboxylation of aminoimidazole ribonucleotide (AIR)[1] to N^5-carboxyaminoimidazole ribonucleotide (N^5-CAIR) as indicated in Scheme 1 (*2*, *3*). Recently, the X-ray crystallographic structure of PurK from *Escherichia coli* was determined to 2.5 Å resolution, and both amino acid sequence and three-dimensional structural analyses indicate that PurK belongs to a growing class of proteins displaying the so-called ATP-grasp fold (*4*). Other members of this superfamily include biotin carboxylase (*5*), carbamoyl phosphate synthetase (*6*), glutathione synthetase (*7*), succinyl-CoA synthetase (*8*), and D-Ala-D-Ala ligase (*9*), among others. All of the above-mentioned enzymes appear to produce a phosphoanhydride intermediate for subsequent attack by a wide range of nucleophiles (*10*). Additionally, the ATP-grasp proteins have a similar molecular architecture consisting of three motifs termed the A-, B-, and C-domains. Many of the enzymes in this superfamily have been well studied, and detailed kinetic and mechanistic information is available. However, while crystal structures of the above-mentioned proteins are known, a complete three-dimensional understanding of the structure−function relationships within

Scheme 1



† This research was supported in part by grants from the NIH (Grant GM55513 to H.M.H. and Grant GM24129 to S.J.B.).

‡ X-ray coordinates have been deposited in the Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ, and will be released upon publication (1EYZ and 1EZ1).

* To whom correspondence should be addressed. E-mail: Hazel_Holden@biochem.wisc.edu. Phone: (608) 262-4988. Fax: (608) 262-1319.

§ University of Wisconsin.

∥ The Pennsylvania State University.

⊥ Dyax Corp.

[1] Abbreviations: AIR, aminoimidazole ribonucleotide; AMPPNP, 5′-adenylyl imidodiphosphate; N^5-CAIR, N^5-carboxyaminoimidazole ribonucleotide; GAR, glycinamide ribonucleotide; HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; MOPS, 3-(*N*-morpholino)propanesulfonic acid; PMSF, phenylmethylsulfonyl fluoride; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

10.1021/bi000926j CCC: $19.00 © 2000 American Chemical Society
Published on Web 07/07/2000

the ATP-grasp superfamily is still lacking due to the fact that X-ray diffraction quality crystals of the various enzymes complexed with multiple substrates have not been available to date.

In recent years, additional members of the ATP-grasp superfamily have been identified on the basis of amino acid sequence alignments. One such protein, the *E. coli* PurT-encoded glycinamide ribonucleotide transformylase, or PurT transformylase, was identified on the basis of its degree of amino acid sequence similarity to PurK (27%) (*11*). PurT transformylase catalyzes the formylation of glycinamide ribonucleotide or GAR using ATP and formate as indicated in Scheme 1. Biochemical studies have indicated that PurT transformylase is unable to utilize formyl tetrahydrofolate and therefore differs from the PurN-encoded GAR transformylase, which also produces formyl GAR (*12*). As in other ATP-grasp enzymes, the catalytic mechanism of PurT transformylase appears to proceed through an acyl phosphate intermediate (*12*). Both positional isotope exchange studies and the fact that purified formyl phosphate is chemically and kinetically competent suggest that the acyl phosphate intermediate in the PurT transformylase reaction is formyl phosphate rather than the more common carboxyphosphate intermediate believed to occur in biotin carboxylase, carbamoyl phosphate synthetase, and PurK, for example (*12*). Also consistent with a phosphoanhydride intermediate is the fact that ^{18}O from formate ends up in the phosphate produced in the reaction. PurT transformylase, however, is found only in a few organisms, and the selective pressure to maintain this enzyme in the genome is unclear. Strikingly, however, another enzyme involved in purine metabolism, the PurU-encoded formyltetrahydrofolate hydrolase, appears to act in concert with PurT transformylase by producing the required formate (*13*).

Both the similarities in the amino acid sequences and the reactions catalyzed by PurT transformylase and PurK suggest that additional structural analyses on PurT transformylase may provide valuable new information regarding the structural basis of catalysis as well as the evolutionary relationships within the ATP-grasp superfamily. In attempt to further address the various structure−function issues regarding the ATP-grasp superfamily, such as the manner in which the acyl phosphate intermediate is stabilized, we initiated a high-resolution X-ray crystallographic analysis of the PurT transformylase from *E. coli*. Here we describe the three-dimensional structures of PurT transformylase complexed with the nonhydrolyzable ATP analogue AMPPNP (*14*) and complexed with both AMPPNP and GAR. The latter complex is of special significance in that it represents the first glimpse of a ternary complex for any enzyme belonging to the ATP-grasp superfamily involved in purine biosynthesis.

MATERIALS AND METHODS

Purification of PurT Transformylase. Recombinant PurT transformylase was expressed in an *E. coli* BL21(DE3) overproducing strain, pLysS, carrying the PurT gene on a pET-22b plasmid. All purification procedures were carried out at 4 °C. A 6 L growth typically yielded 30 g of cells that were harvested by centrifugation at 5000 rpm for 15 min using a JA14 rotor. These cells were frozen in liquid nitrogen and stored at −80 °C until they were needed.

For purification, typically 30 g of cells were resuspended and thawed in 100 mL of buffer containing 50 mM HEPES (pH 8.0), 17 mg of PMSF, 0.4 mg of leupeptin, and 0.7 mg of pepstatin. The thawed cells were disrupted by sonication, and the cellular debris was removed by centrifugation at 14 000 rpm for 30 min using a JA14 rotor. A 10% (w/v) solution of streptomycin sulfate was subsequently added (1 mL per 10 mL of supernatant) and the mixture allowed to stir for 30 min. Precipitated material was removed by centrifugation at 14 000 rpm for 30 min using a JA20 rotor. The resulting supernatant was diluted to 300 mL and spun down a final time at 20 000 rpm (JA20 rotor) for 45 min.

The supernatant was loaded onto a 5.0 cm × 25.0 cm fast-flow Q-Sepharose column and washed with 50 mM HEPES (pH 8.0) until the flow-through contained minimal protein. This required approximately 1 L of wash buffer. The protein was then eluted with a 3.0 L (0.0 to 0.50 M) KCl gradient. Fractions containing significant amounts of PurT transformylase, as determined by SDS−PAGE, were pooled, concentrated to 55 mg/mL, and dialyzed overnight against 4 L of buffer containing 50 mM HEPES (pH 8.0) and 50 mM NaCl.

The dialyzed protein was clarified by centrifugation at 20 000 rpm (JA20 rotor) for 45 min. The protein was further purified using a 2.6 cm × 60 cm Pharmacia Superdex-200 sizing column (6 mL load volume). Fractions were pooled on the basis of purity, as determined by SDS−PAGE, dialyzed against a solution containing 15 mM HEPES (pH 7.5) and 50 mM NaCl, and concentrated to 22 mg/mL before being aliquoted and frozen in liquid nitrogen. The total yield from 30 g of cells was approximately 500 mg of PurT transformylase.

Crystallization of PurT Transformylase. The protein stock solution, at 22 mg/mL, was adjusted with AMPPNP, $MgCl_2$, and NaCl such that the final concentrations were 5, 10, and 250 mM, respectively, and the protein concentration was 20 mg/mL. Crystals were grown by the hanging drop method of vapor diffusion at 4 °C. The protein solution was mixed 1:1 with a precipitant solution containing 18−22% (w/v) methyl ether poly(ethylene glycol) 5000, 100 mM MOPS (pH 6.7), and 100 mM $MgCl_2$. After several days, an oily precipitate appeared from which large irregularly shaped crystals developed after 2−3 weeks. Some of these crystals achieved maximum dimensions of 5.0 mm × 3.0 mm × 1.0 mm. The crystals belonged to space group $P2_12_12$ with one dimer per asymmetric unit and the following unit cell dimensions: $a = 62.3$ Å, $b = 179.5$ Å, and $c = 75.7$ Å.

X-ray Data Collection and Processing. X-ray data were collected with a HiStar (Bruker AXS) area detector system using CuKα radiation generated from a Rigaku RU200 rotating anode generator operated at 50 kV and 90 mA and equipped with Göbel focusing optics. The X-ray data sets were processed with SAINT (Bruker AXS, Inc.) and internally scaled with XSCALIBRE (G. Wesenberg and I. Rayment, unpublished results). An initial native X-ray data set was collected at 0 °C to 1.9 Å resolution. Two isomorphous heavy atom derivatives were subsequently prepared by soaking crystals in 1.0 mM methylmercury acetate for 1.5 days or 5.0 mM potassium gold(I) cyanide for 3 days. X-ray data sets for the mercury and gold heavy atom derivatives were collected to 1.9 and 2.2 Å, respectively. The *R*-factors between the native and the mercury and gold heavy atom derivative data sets were 21.2% and

Table 1: X-ray Data Collection Statistics

data set	resolution (Å)	no. of independent reflections	completeness (%)	redundancy	avg I/avg $\sigma(I)$	R_{sym}[a]
enzyme−AMPPNP complex						
native	30.0−1.90	68412	96.8	3.0	12.1	5.6
highest-resolution bin	1.99−1.90	7416	82.2	1.6	2.3	18.6
methyl-HgOAc	30.0−1.90	64310	90.1	2.3	10.3	7.6
highest-resolution bin	1.99−1.90	5636	62.5	1.2	1.8	21.7
$KAu(CN)_2$	30.0−2.20	44218	96.4	2.6	9.2	8.8
highest-resolution bin	2.30−2.20	5105	90.6	1.5	2.5	21.8
enzyme−AMPPNP complex						
native	30.0−1.75	83063	96.1	3.2	13.0	6.7
highest-resolution bin	1.83−1.75	9709	90.8	2.3	2.6	26.9
enzyme−AMPPNP−GAR complex						
native	30.0−1.75	83477	96.6	3.8	20.6	4.0
highest-resolution bin	1.83−1.75	9995	93.5	3.1	5.9	14.0

[a] $R_{sym} = (\Sigma|I - \bar{I}|/\Sigma I) \times 100$.

Table 2: Relevant Least-Squares Refinement Statistics

	enzyme−AMPPNP complex	enzyme−AMPPNP−GAR complex
resolution limits (Å)	30.0−1.75	30.0−1.75
R-factor[a] (overall) (%)/no. of reflections	18.7/83396	18.9/82946
R-factor (working) (%)/no. of reflections	18.5/75056	18.9/74651
R-factor (free) (%)/no. of reflections	22.8/8340	25.4/8295
no. of protein atoms	5949[b]	5948[c]
no. of heteroatoms	929[d]	1044[e]
weighted root-mean-square deviations from ideality		
bond lengths (Å)	0.011	0.012
bond angles (deg)	2.21	2.42
trigonal planes (Å)	0.006	0.008
general planes (Å)	0.011	0.014
torsional angles (deg)	17.5	17.1

[a] R-factor $= (\Sigma|F_o - F_c|/\Sigma|F_o|) \times 100$, where F_o is the observed structure factor amplitude and F_c is the calculated structure factor amplitude. [b] These include multiple conformations for Q 93 in subunit I and R 43 and R 380 in subunit II. [c] These include multiple conformations for Q 126, D 227, and S 361 in subunit I and R 316 in subunit II. [d] These include 846 water molecules, two AMPPNP molecules, five magnesium ions, two sodium ions, one chloride ion, and one MOPS molecule. [e] These include 919 water molecules, two AMPPNP molecules, two GAR molecules, four magnesium ions, two sodium ions, two acetate anions, and 1 MOPS molecule.

17.6%, respectively [where $R = (\Sigma|F_N - F_H|/\Sigma|F_N|) \times 100$ where $|F_N|$ is the native structure factor amplitude and $|F_H|$ is the derivative structure factor amplitude]. Relevant intensity statistics can be found in Table 1.

The heavy atom binding sites were located and their positions refined by least-squares analysis with the software package SOLVE (*15*, *16*). The heavy atom derivatives were placed on a common origin by appropriate difference Fourier maps. Protein phases were calculated with SOLVE. The initial electron density map was subsequently improved via cyclical averaging and solvent flattening as implemented in DM (*17*). From this map, it was possible to trace the polypeptide chain from Thr 2 to Gly 392 with only one break in the electron density between Ala 187 and Gly 190.

A final high-resolution native X-ray data set was collected to 1.75 Å on a flash-frozen crystal that had been transferred to a cryoprotectant solution containing 2.5 mM AMPPNP, 500 mM NaCl, 25% (w/v) methyl ether poly(ethylene glycol) 5000, 50 mM $MgCl_2$, 100 mM MOPS (pH 6.7), and 15% (v/v) ethylene glycol. The structure was refined by least-squares analysis with the software package TNT (*18*). Relevant refinement statistics can be found in Table 2. The average B-values for the two subunits in the asymmetric unit, the nucleotides, and the solvents were 28.0, 23.4, and 37.4 Å^2, respectively. All backbone atoms for the two subunits in the asymmetric unit superimpose with a root-mean-square deviation of 0.54 Å. A portion of the electron density map near the nucleotide is displayed in Figure 1.

For preparation of the complex with GAR, crystals grown in the presence of AMPPNP were transferred to a solution containing 20% (w/v) methyl ether poly(ethylene glycol) 5000, 50 mM $MgCl_2$, 350 mM NaCl, 2.5 mM AMPPNP, 100 mM MOPS (pH 6.7), and 1 mM GAR. The crystals were allowed to soak for 12 h, after which they were transferred to a cryoprotectant solution in which the poly(ethylene glycol) concentration had been increased to 25% (w/v) and the NaCl concentration to 500 mM. Additionally, 1 M sodium acetate was included in the cryoprotectant solution. An X-ray data set to 1.75 Å resolution was collected from a single crystal in the same manner as that described for the PurT transformylase−AMPPNP complex. The structure was determined by difference Fourier analysis and refined with the program TNT. Relevant intensity and refinement statistics can be found in Tables 1 and 2, respectively. The average B-values for all polypeptide chain backbone atoms for subunits I and II were 16.1 and 31.2 Å^2, respectively. In subunit II, the electron density for the polypeptide chain from Glu 123 to Gly 196 was less well-ordered than that for the corresponding region in subunit I, and this difference is most likely due to crystalline packing. A portion of the electron density map near the GAR moiety is displayed in Figure 2.

RESULTS AND DISCUSSION

Molecular Structure of the PurT Transformylase−AMPPNP Complex. As shown in Figure 3, the PurT transform-

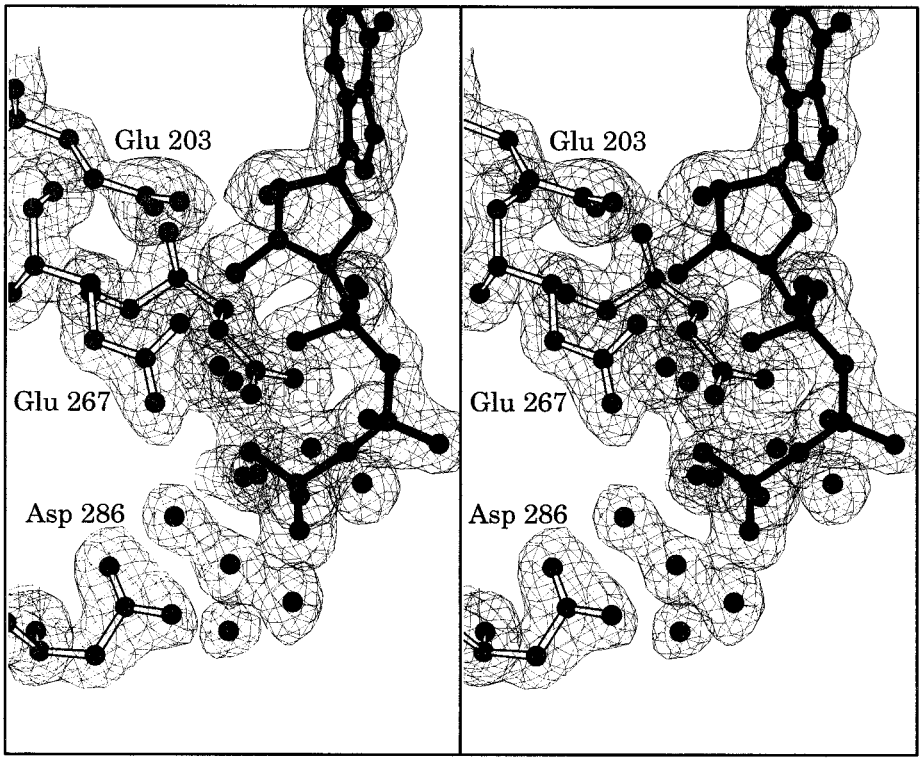


FIGURE 1: Region of the electron density map corresponding to the bound AMPPNP molecule. The map was contoured at 1σ and calculated with coefficients of the form $2F_o - F_c$, where F_o was the native structure factor amplitude and F_c was the calculated structure factor amplitude. This figure was produced with the software package BobScript (*26*).

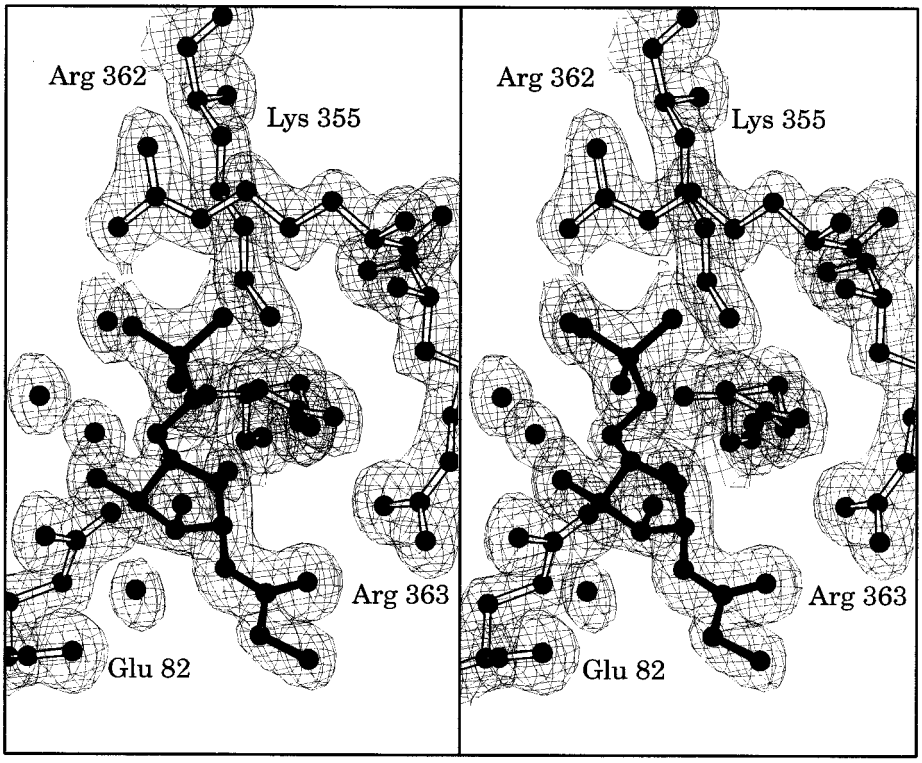


FIGURE 2: Region of the electron density map corresponding to the bound GAR moiety. The map was calculated and contoured as described in the legend of Figure 1. This figure was produced with the software package BobScript (*26*).

ylase is a symmetrical dimer with overall molecular dimensions of approximately 68 Å × 92 Å × 79 Å. Each subunit contains 392 amino acid residues and can be described in terms of three structural motifs, the A-, B-, and C-domains which are color-coded in Figure 3 in yellow, green, and cyan, respectively. The A-domain is formed by Thr 2−Ala 122 and is dominated by a five-stranded parallel β-pleated sheet flanked on either side by two α-helices. The β-strands range in length from four to six amino acid residues, while the α-helices range in length from four to eleven amino acid

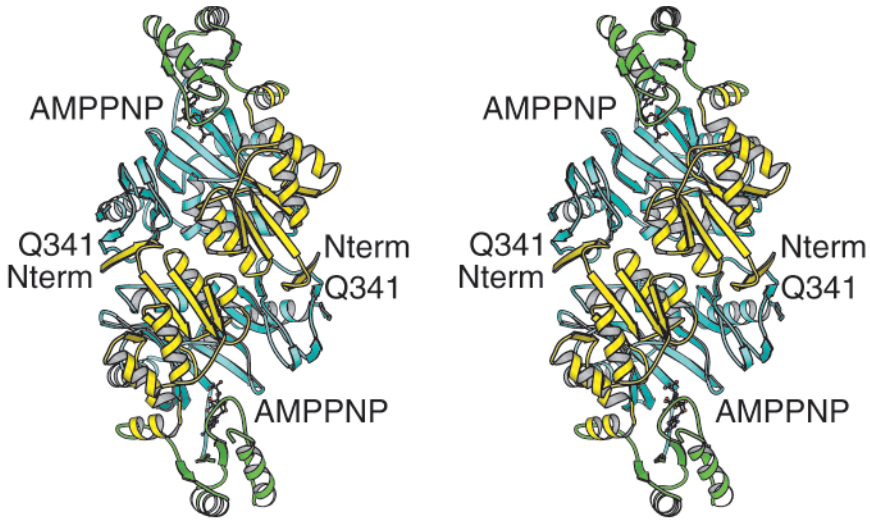


FIGURE 3: Ribbon representation of the PurT transformylase dimer. The 2-fold rotational axis of the dimer is oriented perpendicular to the plane of the paper. Each subunit can be described in terms of three structural motifs, the A-, B-, and C-domains which correspond to Thr 2−Ala 122, Glu 123−Gly 196, and Val 197−Gly 392 and color-coded in yellow, green, and cyan, respectively. All ribbon representations were prepared with the software package MOLSCRIPT (*26*).

residues. Connecting these secondary structural elements are numerous turns, including four type I and two type II classical reverse turns. Leading from the A-domain into the B-domain is a stretch of polypeptide chain delineated by Ala 105−Ala 122 that adopts a helix−residue−helix motif. The dihedral angles for the intervening residue, namely, Asn 113 ($\phi = -138°$, $\psi = 129°$), are such that the axes of the two α-helices lie at nearly right angles to one another. This tertiary pattern is a hallmark for enzymes belonging to the ATP-grasp superfamily (*4*).

The smallest of the three structural motifs of the PurT transformylase, the B-domain is formed by Glu 123−Gly 196 and contains a four-stranded antiparallel β-sheet with the strands ranging in length from three to five amino acid residues. One side of the β-sheet is covered by two α-helices delineated by Glu 138−Ile 148 and Leu 174−Gln 183. In addition to these secondary structural elements, there is one type I turn, one type I′ turn, and one type II turn. A break in the electron density for the polypeptide chain occurs between Ala 187 and Gly 190.

Clearly, the most complicated of the domains, the C-motif extends from Val 197 to Gly 392 and is composed primarily of an eight-stranded antiparallel β-pleated sheet formed by Phe 202−Ser 210, Val 215−Gln 225, Tyr 230−Gln 235, Gly 262−Val 270, Val 275−Ser 281, Ala 321−Ile 326, Gln 349−Leu 352, and Gly 365−Thr 370. There are two additional regions of antiparallel β-sheet formed by Gln 329−Ser 332 and Ile 358−Ser 361 and Thr 336−Asp 338 and Val 388−Gly 392, respectively. The four α-helices located in the C-domain range in length from four to eighteen amino acid residues. There are seven classical reverse turns in the C-domain that link these various β-strands and α-helices together (three type I, one type I′, one type II, one type II′, and one type III).

The dimeric interface is formed from regions provided by both the A- and C-domains. Specifically, in the A-domains, there are three major regions that contribute to the subunit−subunit interface: the extended region at the N-terminus (Thr 2−Ala 12), the first α-helix (Glu 22−Cys 31) which ends in a type I turn (Gln 32−Gly 35), and the second α-helix

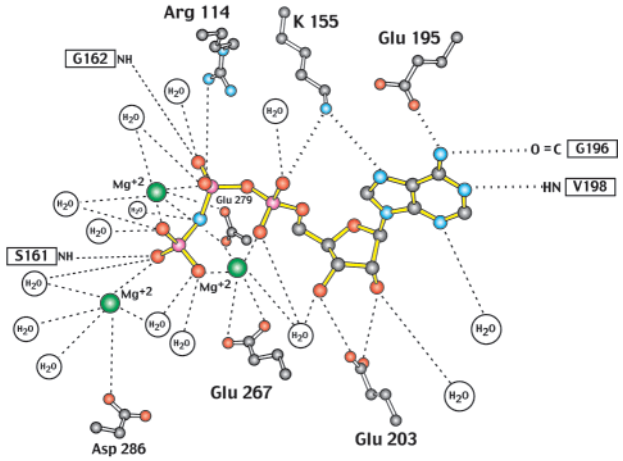


FIGURE 4: Cartoon of the hydrogen bonding pattern surrounding the AMPPNP moiety. Potential hydrogen bonds, within 3.1 Å, are represented by the dashed lines. All three magnesium ions, represented as green spheres, are octahedrally coordinated by oxygen ligands.

(Pro 48−His 51). Indeed, as can be seen in Figure 3, the N-terminus of subunit I participates in an antiparallel β-sheet with residues Val 335−Gln 341 of subunit II. The region of polypeptide chain in the C-domain that is primarily involved in the formation of the subunit−subunit interstitial region is delineated by Gln 333−Lys 355 and includes two strands of β-sheet (Thr 336−Asp 338 and Gln 349−Leu 352), one type III turn (Asn 339−Asn 342), one type I turn (Val 340−Ala 343), and one type II turn (Gly 345−Leu 348). The surface area buried upon dimerization is approximately 5000 Å^2 as calculated according to the algorithm of Lee and Richards (*19*) with a probe radius of 1.4 Å.

The binding pockets for the AMPPNP moieties are wedged between the B- and C-domains and are separated by approximately 50 Å in the dimer. A cartoon representation of the hydrogen bonding pattern around the AMPPNP moiety in subunit I is displayed in Figure 4. Approximately 90% of the surface area of the AMPPNP moiety is buried upon binding to PurT transformylase. Twelve water molecules lie within hydrogen bonding distance of the purine nucleotide.

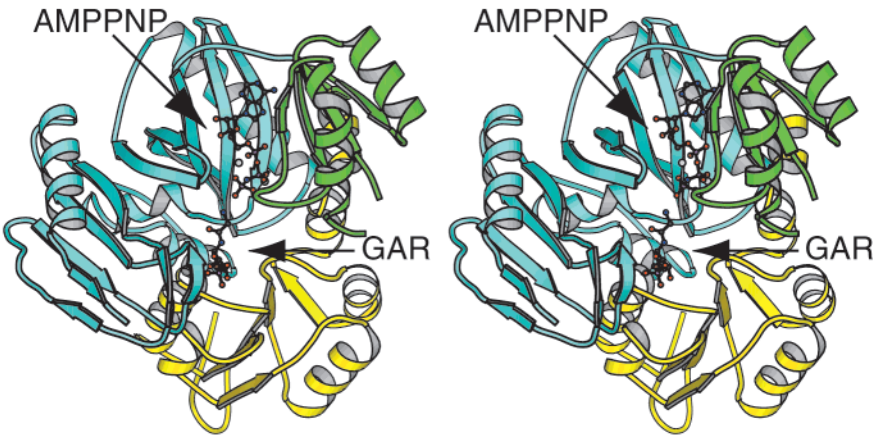


FIGURE 5: Ribbon representation of the PurT transformylase subunit. The AMPPNP and GAR moieties are depicted as a ball-and-stick representation. One of the γ-phosphoryl oxygens of the AMPPNP is positioned at 2.8 Å from the amino group of GAR.

The adenine ring is anchored to the protein via the ϵ-amino group of Lys 155, the carboxylate group of Glu 195, and the peptidic carbonyl and NH groups of Gly 196 and Val 198, respectively. The ribose of the AMPPNP moiety adopts the $C_{3'}$-*endo* pucker. As observed in other ATP-grasp proteins such as carbamoyl phosphate synthetase (*20*), the 2′- and 3′-hydroxyl groups of the purine ribose lie within hydrogen bonding distance of a glutamate residue, namely, Glu 203 in PurT transformylase. In carbamoyl phosphate synthetase, the equivalent glutamate residue serves both to bridge the 2′- and 3′-hydroxyl groups of the AMPPNP ribose and to act as a ligand to a physiologically important potassium ion (*20*). In PurT transformylase, Arg 114 and Lys 155 interact with the β- and α-phosphoryl moieties of the AMPPNP, respectively. Arg 114 lies within the helix−residue−helix motif leading from the A-domain into the B-domain. The peptidic NH groups of Ser 161 and Gly 162 are positioned within hydrogen bonding distance of a γ- and a β-phosphoryl oxygen, respectively. The structurally equivalent residues in carbamoyl phosphate synthetase, Gly 721 and Gly 722, have been implicated as the trigger for the movement of the B-domain over the active site upon nucleotide binding (*20*). There are three magnesium ions located within the nucleotide binding pocket of subunit I in the asymmetric unit. Each of these metal ions is surrounded in an octahedral coordination sphere by oxygen-containing ligands. The average ligand−metal bond length is 2.2 Å. One of the magnesium ions is coordinated by an α- and a γ-phosphoryl oxygen, while the second is wedged between a β- and a γ-phosphoryl oxygen. These two magnesium ions are located in positions nearly identical to those found in carbamoyl phosphate synthetase (*20*). Like that observed in carbamoyl phosphate synthetase, a glutamate residue, Glu 279 in PurT transformylase, bridges these magnesium ions. Additionally, as in carbamoyl phosphate synthetase, a glutamate residue, Glu 267 in PurT transformylase, serves as a bidentate ligand to one of the magnesiums (Figure 4). The third magnesium ion in PurT transformylase is situated at the end of the AMPPNP ligand and ligated by a γ-phosphoryl oxygen of the nucleotide, the carboxylate group of Asp 286, and four solvent molecules. This magnesium ion is not observed in subunit II of the dimer.

Molecular Structure of the PurT−AMPPNP−GAR Complex. As in the PurT transformylase−AMPPNP complex, crystals of PurT transformylase complexed with AMPPNP and GAR belonged to space group $P2_12_12$ and contained one dimer per asymmetric unit. In that the electron density corresponding to subunit I of the dimer was much better defined, the following discussion refers only to this monomer. Shown in Figure 5 is a ribbon representation of subunit I with bound AMPPNP and GAR. The monomer has overall molecular dimensions of 55 Å × 69 Å × 44 Å. As can be seen, GAR sits in a binding pocket situated at the apex of the A-, B-, and C-domains. Approximately 75% of the surface area for GAR becomes buried upon binding to PurT transformylase.

A closeup view of the active site for PurT transformylase is shown in Figure 6a. The binding pocket is fairly hydrophilic with twenty ordered water molecules lying within 3.2 Å of the GAR and AMPPNP moieties. The amino group of GAR is positioned at 2.8 Å from one of the γ-phosphoryl oxygens of AMPPNP. The third magnesium ion observed in the PurT transformylase−AMPPNP complex is displaced upon GAR binding. Indeed, the carboxylate group of Asp 286 which acts as a ligand to the third magnesium in the PurT transformylase−AMPPNP complex now hydrogen bonds to the amino group of GAR as indicated in Figure 6b. Glu 82 serves to bridge the 2′- and 3′-hydroxyl groups of the GAR ribose. Like that observed with AMPPNP, the ribose of GAR adopts the $C_{3'}$-*endo* pucker. Three positively charged amino acid residues involved in GAR binding are Lys 355, Arg 362, and Arg 363 as indicated in Figure 6b. Specifically, the guanidinium group of Arg 363 interacts with the carbonyl oxygen of the glycinyl moiety of GAR, while the side chains of Lys 355 and Arg 362 lie within hydrogen bonding distance of the phosphoryl group of the substrate. There are nine solvent molecules that hydrogen bond to GAR, seven of which surround the phosphoryl group.

The GAR substrate is positioned at the C-terminal end of the five-stranded parallel β-sheet of the A-domain (Figure 5). As shown in Figure 7, this β-sheet is topologically equivalent to the Rossmann fold observed, for example, in the NAD-dependent dehydrogenases and in UDP-galactose 4-epimerase (*21*). Indeed, the α-carbons for the A-domain of PurT transformylase and the N-terminal motif of UDP-galactose 4-epimerase superimpose with a root-mean-square deviation of 0.9 Å for 53 structurally equivalent atoms. The degree of amino acid sequence similarity between these two Rossmann folds is ∼44%. Strikingly, the ribose of GAR in PurT transformylase and the diphosphate moiety of NADH

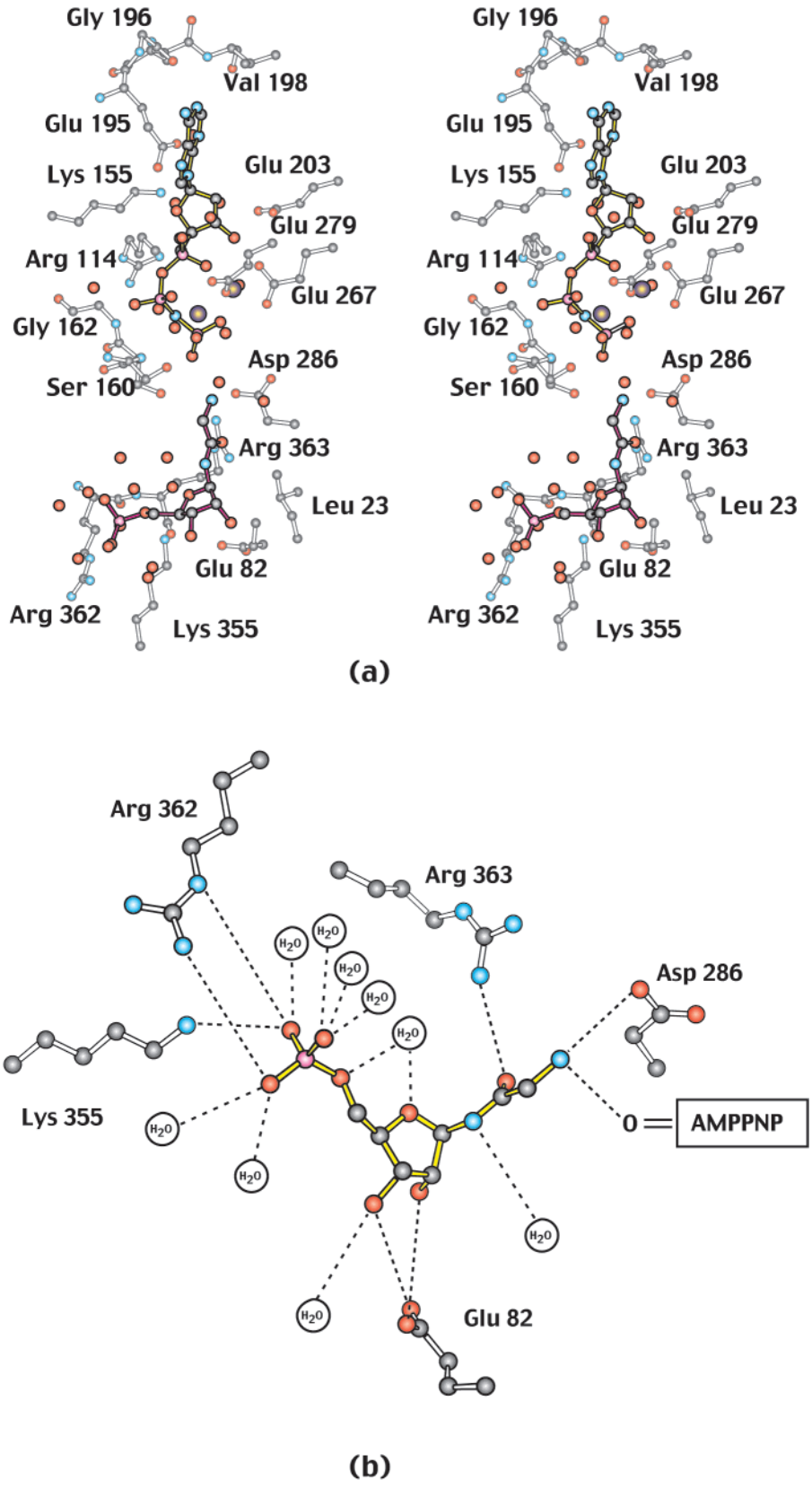


FIGURE 6: Active site of PurT transformylase with bound AMPPNP and GAR. Those amino acid residues lying within 3.2 Å of the AMPPNP and GAR molecules are shown in panel a. Potential hydrogen bonds (within 3.1 Å) between GAR and the protein are represented by the dashed lines in panel b.

in the epimerase occupy similar positions with respect to the Rossmann fold.

Comparison of PurT Transformylase with PurK. As discussed in the introductory section, the three-dimensional structure of PurK was recently determined to 2.5 Å resolution in the presence of Mg^{2+}ADP and to 2.1 Å resolution in a sulfate-liganded state (*4*). Recall from Scheme 1 that PurK functions in purine biosynthesis by catalyzing the conversion

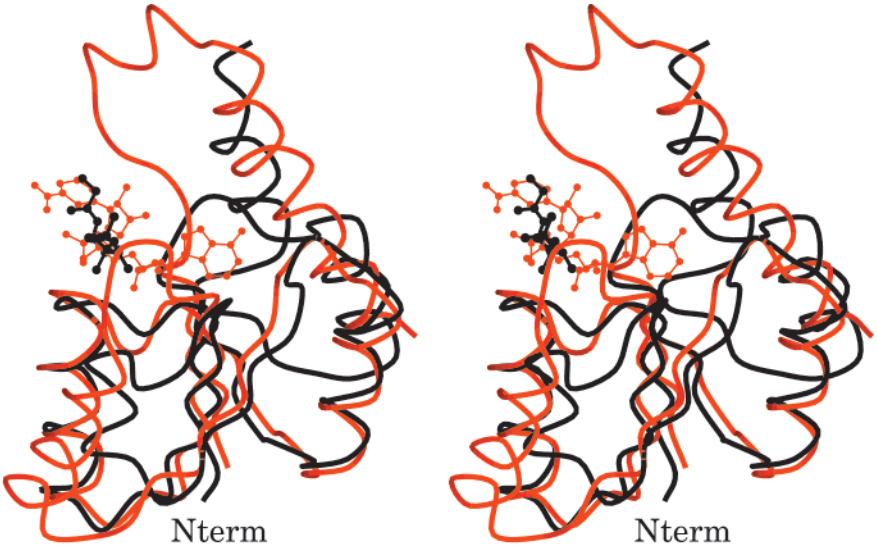


FIGURE 7: Superposition of the Rossmann folds found in UDP-galactose 4-epimerase and PurT transformylase. The course of the polypeptide chain for UDP-galactose 4-epimerase is highlighted in red, while that for PurT transformylase is given in black. The NADH and GAR molecules are shown in ball-and-stick representations. Note the relative orientation of GAR with respect to NADH.

of 5-aminoimidazole ribonucleotide (AIR) to 4-carboxy-5-aminoimidazole ribonucleotide (N^5-CAIR). Like PurT transformylase, PurK has a catalytic mechanism thought to proceed through an acyl phosphate intermediate. The amino acid sequences for these two enzymes are 55% homologous with 27% of the residues being identical. It has been postulated that the sulfate binding site for the sulfate-liganded form of PurK may be the region of polypeptide chain responsible for stabilization of the carboxyphosphate intermediate (*4*).

A superposition of the α-carbon traces for the PurT transformylase−AMPPNP−GAR complex and the PurK−Mg^{2+}ADP complex is displayed in Figure 8a. These two molecules superimpose with a root-mean-square deviation of 1.1 Å for 177 structurally equivalent α-carbons. As can be seen, the nucleotides for the two proteins are accommodated within their respective binding pockets in similar manners. While the structure of PurK was determined only in the presence of bound Mg^{2+}ADP, it can be speculated, however, that the ribotide substrates for PurK and PurT transformylase, namely, AIR and GAR, respectively, might bind in similar but perhaps not identical manners. With this caveat in mind, it can be predicted that the hydroxyl groups of the AIR ribose are anchored to PurK via Glu 49, which is the structural equivalent of Glu 82 in PurT transformylase. Likewise, as indicated in Figure 6b, Lys 355 and Arg 362 in PurT transformylase interact with the phosphoryl group of GAR, and we can speculate that the structurally equivalent residues in PurK, namely, Lys 307 and Arg 313, play a similar role in AIR binding.

Formate Binding Pocket. A number of enzymes of the ATP-grasp superfamily, most notably biotin carboxylase, carbamoyl phosphate synthetase, PurK, and PurT transformylase, have evolved mechanisms for recognizing and binding small carboxylic acid substrates. Despite X-ray crystallographic structures for all of these enzymes, the nature of the carboxylic acid binding site in these proteins has yet to be elucidated. For PurT transformylase, the formate binding pocket must, a priori, be located near the site of chemistry, namely, the γ-phosphate of ATP. Inspection of the region surrounding the terminal phosphate indicates a paucity of amino acids that could form hydrogen bonds with formate (Figure 6a). One clue to the location of the formate binding site comes from the sulfate-liganded structure of PurK. In PurK, Arg 242 forms a salt bridge to the sulfate, which is thought to mimic the location of the carboxyphosphate intermediate in PurK (Figure 8b). Arg 242 is conserved among all PurT transformylases and PurK enzymes, suggesting a potential role in catalysis and carboxylic acid binding. Interestingly, another formate utilizing enzyme, formate dehydrogenase, utilizes an arginine to form the binding pocket for formate (*22*), and the enzyme D-alanine-D-alanine ligase is thought to utilize Arg 255 to bind to the carboxylate of D-alanine (*23*). In the structure of PurT transformylase, Arg 283, equivalent to Arg 242 in PurK, adopts a conformation significantly different from that observed in the PurK enzyme (Figure 8b). However, Arg 283 could adopt a different conformation upon formate binding. Site-directed mutagenesis studies have indeed demonstrated that the R283A mutation produces a completely inactive enzyme (*24*). However, a catalytic role for this residue could not be definitively determined since the protein displayed abnormal expression and solubility characteristics indicating that this mutation also affected the overall structure of the protein.

A second clue to the location of the formate binding site comes from site-directed mutagenesis studies on PurT transformylase. One mutation, G162I, was prepared on the basis of amino acid sequence alignments with other members of the ATP-grasp superfamily. Kinetic studies on this enzyme indicated a 317-fold reduction in k_{cat}, with a 2-fold increase in K_m for ATP and GAR (*12*). Most strikingly, G162I displayed a 180-fold increase in K_m for formate, suggesting a direct effect on the formate binding pocket. Examination of the ternary complex of PurT in Figure 6a indicates that Gly 162 occurs as part of a surface loop and is positioned along one face of the nucleotide ~6 Å away from the γ-phosphate of ATP. Given this three-dimensional constraint, it is difficult to propose a direct interaction between Gly 162 and formate. However, the G162I mutation should perturb

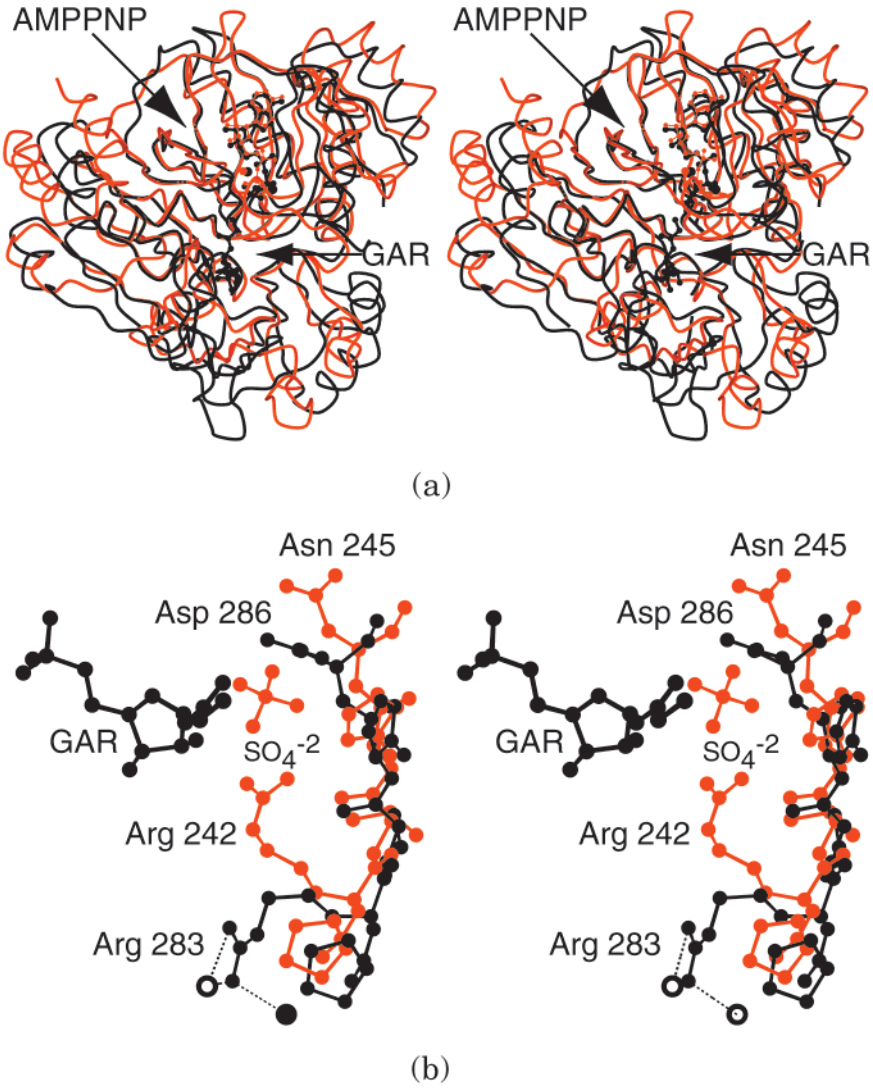


FIGURE 8: Comparison of PurT transformylase with PurK. A superposition of the α-carbons for PurT transformylase (in black) and PurK (in red) is given in panel a. The structure of PurK employed for this particular superposition was determined in the presence of Mg^{2+}ADP as indicated by the ball-and-stick representation. The positions of GAR and AMPPNP in PurT transformylase are depicted in black balls and sticks. A second structure of PurK was determined with only a sulfate bound in the active site. The region of polypeptide chain surrounding this sulfate in PurK is shown in panel b in red filled bonds. Superimposed upon this region is the corresponding area in the PurT transformylase−AMPPNP−GAR complex depicted in black filled bonds. It is postulated that these regions of the proteins may be responsible for stabilizing the acyl phosphate intermediates that occur in reactions catalyzed by members of the ATP-grasp superfamily.

the location of the γ-phosphate and the Mg^{2+} ions in the active site as well as disrupt the relative orientations of the GAR amine with respect to ATP. The mechanism by which this perturbation is accomplished is not known; however, the most likely explanation is that the Gly 162 to Ile mutation results in a surface loop that does not close tightly around ATP, thereby resulting in improper positioning of the nucleotide within the binding pocket. Such a domain movement has been observed for PurK upon binding of Mg^{2+}ADP to the active site (*4*). As a consequence of this "open" conformation, G162I PurT transformylase releases formyl phosphate whereas wild-type PurT transformylase has no detectable formyl phosphate release.

How does the suboptimal orientation of GAR and ATP affect formate binding? It is proposed that the formate binding pocket is formed by the Mg^{2+} ions, the γ-phosphoryl oxygen, and the GAR amine functional group (Figure 6a). The subtle movement of the ATP relative to GAR in the G162I mutant would destroy the optimal alignment of hydrogen bonding residues and thus increase the K_m for formate. The large decrease in k_{cat} also suggests an improper alignment between the formyl phosphate and the nucleophilic amine of GAR.

Currently, it is not possible to assign those residues responsible for formate binding, nor to outline a detailed hydrogen bonding pattern scheme. Examination of the crystal structure of PurT transformylase indicates that the Mg^{2+} ions are roughly 3.5 Å away from where formate might be expected to bind. Thus, it is hard to envision a direct contact between Mg^{2+} and formate. However, the active site contains a number of water molecules, and some of these could serve as a bridging atom to anchor formate to Mg^{2+}.

Structural Clues to the Mechanism of PurT Transformylase. In an effort to understand some of the molecular features involved in catalysis, the structure of PurT transformylase was compared to that of D-alanine-D-alanine ligase (*9*). This particular comparison was chosen since D-alanine-D-alanine ligase was crystallized in the presence of a mechanism-based inhibitor bound in the active site (*23*). In that D-alanine-D-alanine ligase must act upon this compound, a number of

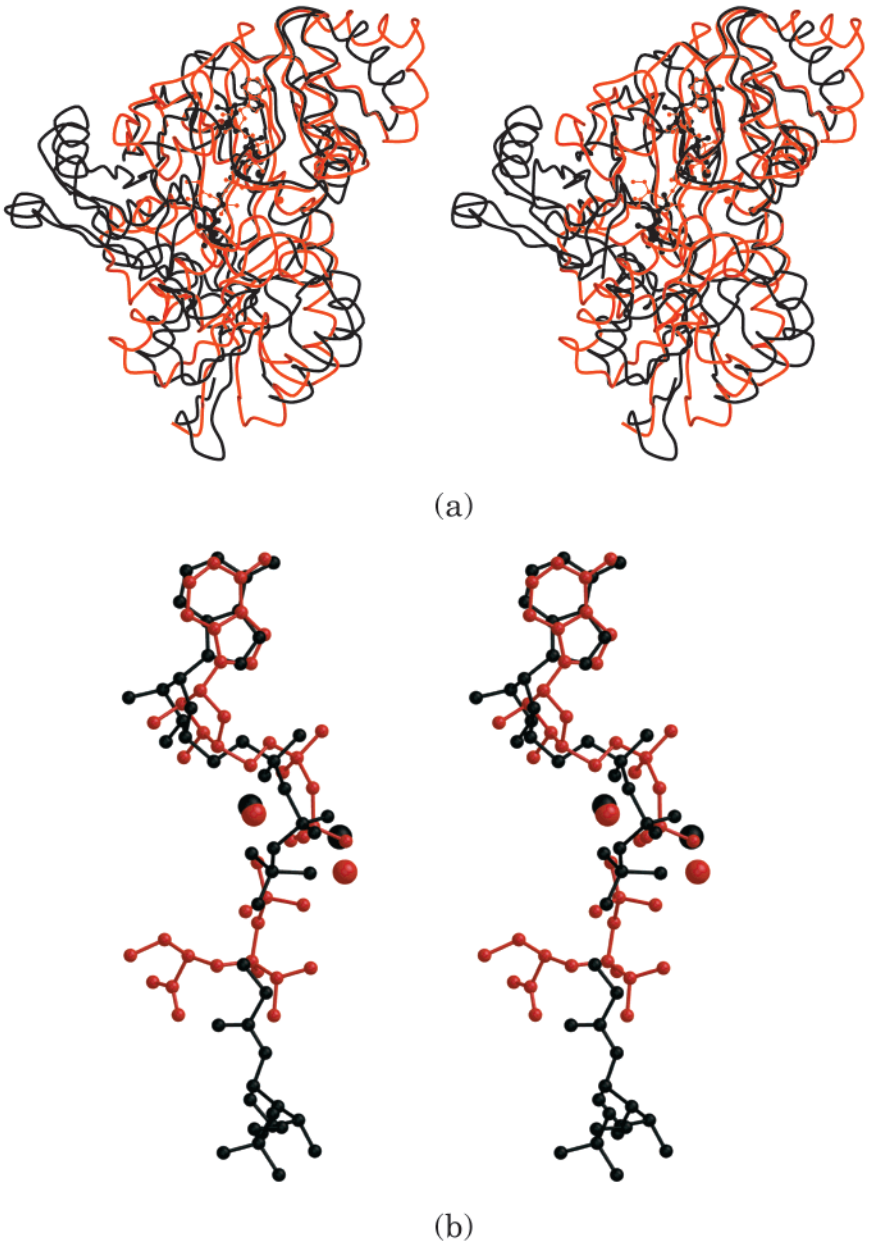


FIGURE 9: Comparison of PurT transformylase with D-alanine-D-alanine ligase. Shown in panel a is a superposition of PurT transformylase, in black, onto D-alanine-D-alanine ligase, in red. An overlay of the ligands for these enzymes is displayed in panel b. The AMPPNP and GAR moieties in PurT transformylase are drawn in black, while the mechanism-based inhibitor of D-alanine-D-alanine ligase is displayed in red.

features of the chemistry catalyzed by D-alanine-D-alanine ligase would be expected to be similar for the inhibitor and the wild-type substrates. The crystal structures of PurT transformylase and D-alanine-D-alanine ligase were superimposed according to the algorithm of Rossmann and Argos (*25*), and using a conserved amino acid sequence in the ATP-grasp domain. As can be seen from Figure 9a, the two structures align well with one another and display a root-mean-square difference of 1.8 Å for 167 structurally equivalent α-carbon atoms.

Examination of the active site ligands of the overlaid proteins shows a remarkable degree of similarity (Figure 9b). The adenine rings of the cofactor along with the α- and *β*-phosphate groups align very well within the two structures. There is a slight shift in the Mg^{2+} ions (0.3 and 1.4 Å) between the two structures, the largest difference resulting from the cleavage and movement of the *γ*-phosphate. This movement would be expected for PurT transformylase as well. The D-alanine portion of the mechanism-based inhibitor occupies a pocket behind the ATP ring and is suggestive of the region of the protein responsible for binding of the carboxylic acid. Interestingly, one of the anchoring contacts to the inhibitor in D-alanine-D-alanine ligase is Arg 255, and this amino acid occurs in a position similar to Arg 283 of PurT transformylase, though in a different conformation.

The comparison of PurT transformylase to D-alanine-D-alanine ligase also suggests amino acids that could be important in binding and catalysis. Both D-alanine and GAR are protonated at physiological pH, indicating that the enzymes must utilize mechanisms to facilitate deprotonation of the amines to achieve catalysis. In D-alanine-D-alanine ligase, the nucleophilic D-alanine is deprotonated by Tyr 216 to facilitate catalysis and Ser 150 and Glu 15 are important for the proper alignment of Tyr 216. Examination of the

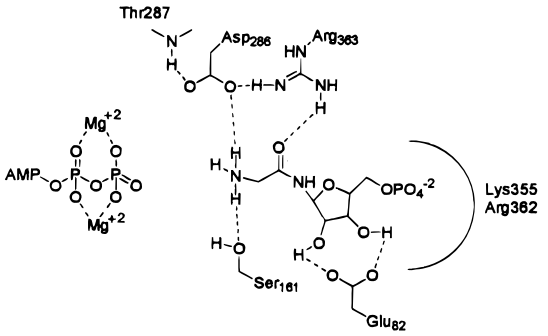


FIGURE 10: Cartoon of the hydrogen bonding pattern around the putative active site base. It is proposed that Asp 268 serves as the active site base to deprotonate the amine group of GAR. Potential hydrogen bonds are represented by the dashed lines.

overlaid structures of PurT transformylase and D-alanine-D-alanine ligase indicates a network of amino acids that could facilitate catalysis by deprotonation of the amine of GAR (Figure 10). Asp 286 is hydrogen bonded to the amine of GAR and could act as the general base in this reaction. The aspartate is held by two additional hydrogen bond contacts, one to Arg 363, which also binds to the carbonyl of GAR, and a backbone amide contact to Thr 287. Further binding of the GAR amine is accomplished by Ser 161, which occupies a position close to that of Ser 150 in D-alanine-D-alanine ligase. These sets of contacts should help to orient the amine of GAR as well as provide the base necessary for its deprotonation.

The most striking feature of the alignment between PurT transformylase and D-alanine-D-alanine ligase comes from a comparison of the amine of GAR and the inhibitor. The mechanism-based inhibitor is formed in the active site by the attack of the oxygen atom of the phosphinate on the γ-phosphate of ATP. The resulting phosphinophosphate intermediate can then be thought of as a mimic of the acyl phosphate intermediate in the reaction. Given this postulate, the phosphinate occupies a position that mechanistically should be similar to the carbonyl carbon atom of the formyl phosphate. Inspection of the overlaid structures in Figure 9a reveals that the amino group of GAR is in a perfect in-line attack on this phosphorus atom. Thus, we can postulate a mechanism in which formate reacts with ATP to form formyl phosphate (Figure 11). The breakage of the $\beta-\gamma$ phosphate bond results in a 1.5 Å shift of the Mg^{2+}-formyl phosphate bringing the carbonyl carbon of the formyl phosphate near the amine of GAR. The amine reacts with the carbonyl group to release phosphate and formyl-GAR.

In summary, the molecular architecture of PurT transformylase has now been defined on the basis of high-resolution X-ray structures of the enzyme complexed with AMPPNP and with AMPPNP and GAR. This study is of special significance in that for the first time the substrate binding pocket for any ATP-grasp enzyme involved in purine metabolism has been identified. From this structure and a comparison of PurT transformylase with other ATP-grasp proteins, it has been possible to outline a possible catalytic mechanism. Several questions remain, however, regarding PurT transformylase. Does formate bind on the enzyme as predicted? How is the formyl phosphate intermediate stabilized? What are the structural relationships within the active site between GAR, formyl phosphate, and Mg^{2+}ADP? Site-

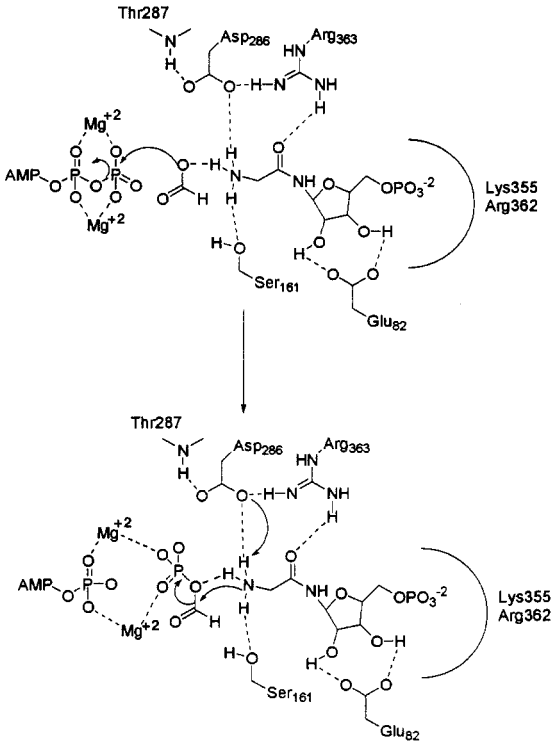


FIGURE 11: Possible catalytic mechanism for PurT transformylase. In this mechanism, the breakage of the $\beta-\gamma$ phosphate bond results in a shift of the Mg^{2+}-formyl phosphate thereby bringing the carbonyl carbon of the formyl phosphate near the amine of GAR. Asp 268 functions to deprotonate the amine of GAR, which then reacts with the carbonyl group of formyl phosphate to release phosphate and formyl-GAR.

directed mutagenesis experiments and additional X-ray crystallographic analyses are in progress in an effort to address these and other biochemical issues regarding both PurT transformylase in particular and the ATP-grasp superfamily in general.

ACKNOWLEDGMENT

We gratefully acknowledge Dr. W. W. Cleland for helpful discussions.

REFERENCES

1. Buchanan, J. M., and Hartman, S. C. (1959) *Adv. Enzymol. Relat. Areas Mol. Biol. 39*, 91−183.
2. Meyer, E., Leonard, N. J., Bhat, B., Stubbe, J., and Smith, J. M. (1992) *Biochemistry 31*, 5022−5032.
3. Mueller, E. J., Meyer, E., Rudolph, H., Davisson, V. J., and Stubbe, J. (1994) *Biochemistry 33*, 2269−2278.
4. Thoden, J. B., Kappock, T. J., Stubbe, J., and Holden, H. M. (1999) *Biochemistry 38*, 15480−15492.
5. Waldrop, G. L., Rayment, I., and Holden, H. M. (1994) *Biochemistry 33*, 10249−10256.
6. Thoden, J. B., Holden, H. M., Wesenberg, G., Raushel, F. M., and Rayment, I. (1997) *Biochemistry 36*, 6305−6316.
7. Yamaguchi, H., Kato, H., Hata, Y., Nishioka, T., Kimura, A., Oda, J., and Katsube, Y. (1993) *J. Mol. Biol. 229*, 1083−1100.
8. Wolodko, W. T., Fraser, M. E., James, M. N. G., and Bridger, W. A. (1994) *J. Biol. Chem. 269*, 10883−10890.
9. Fan, C., Moews, P. C., Walsh, C. T., and Knox, J. R. (1994) *Science 266*, 439−443.
10. Galperin, M. Y., and Koonin, E. V. (1997) *Protein Sci. 6*, 2639−2643.
11. Marolewski, A., Smith, J. M., and Benkovic, S. J. (1994) *Biochemistry 33*, 2531−2537.

12. Marolewski, A., Mattia, K. M., Warren, M. S., and Benkovic, S. J. (1997) *Biochemistry 36*, 6709−6716.
13. Nagy, P. L., Marolewski, A., Benkovic, S. J., and Zalkin, H. (1995) *J. Bacteriol. 177*, 1292−1298.
14. Yount, R. G., Babcock, D., Ballantyne, W., and Ojala, D. (1971) *Biochemistry 10*, 2484−2489.
15. Terwilliger, T. C., and Eisenberg, D. (1983) *Acta Crystallogr. A39*, 813−817.
16. Terwilliger, T. C. (1997) in *Methods in Enzymology* (Carter, C. W., Jr., Sweet, R. M., Abelson, J. N., and Simon, M. I., Eds.) Vol. 276, pp 530−537, Academic Press, New York.
17. Cowtan, K., and Main, P. (1998) *Acta Crystallogr. D54*, 487−493.
18. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr. A43*, 489−501.
19. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
20. Thoden, J. B., Wesenberg, G., Raushel, F. M., and Holden, H. M. (1999) *Biochemistry 38*, 2347−2357.
21. Thoden, J. B., Frey, P. A., and Holden, H. M. (1996) *Protein Sci. 5*, 2149−2161.
22. Lamzin, V. S., Dauter, Z., Popov, V. O., Harutyunyan, E. H., and Wilson, K. S. (1994) *J. Mol. Biol. 236*, 759−785.
23. Fan, C., Park, I.-S., Walsh, C. T., and Knox, J. R. (1997) *Biochemistry 36*, 2531−2538.
24. Marolewski, A. E. (1994) Ph.D. Thesis, The Pennsylvania State University, University Park, PA.
25. Rossmann, M. G., and Argos, P. (1975) *J. Biol. Chem. 250*, 7525−7532.
26. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950 (BobScript, version 2.01, copyright Robert Esnouf, 1994−1996).

BI000926J