A Model for the *Bacillus subtilis* Formylglycinamide Ribonucleotide Amidotransferase Multiprotein Complex[†,‡]

Ruchi Anand,[§] Aaron A. Hoskins,[∥] Eric M. Bennett,[§] Michael D. Sintchak,[∥] JoAnne Stubbe,[⊥] and Steven E. Ealick*[,§]

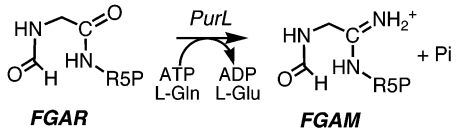
Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, and Departments of Chemistry and Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received April 29, 2004; Revised Manuscript Received June 14, 2004

ABSTRACT: Formylglycinamide ribonucleotide amidotransferase (FGAR-AT) catalyzes the conversion of formylglycinamide ribonucleotide (FGAR), ATP, and glutamine to formylglycinamidine ribonucleotide (FGAM), ADP, P_i, and glutamate in the fourth step of the purine biosynthetic pathway. PurL exists in two forms: large PurL (lgPurL) is a single chain, multidomain enzyme of about 1300 amino acids, whereas small PurL (smPurL) contains about 800 amino acids but requires two additional gene products, PurS and PurQ, for activity. smPurL contains the ATP and FGAR binding sites, PurQ is a glutaminase, and the function of PurS is just now becoming understood. We determined the structure of *Bacillus subtilis* PurS in two different crystal forms $P2_1$ and $C2$ at 2.5 and 2.0 Å resolution, respectively. PurS forms a tight dimer with a central six-stranded β-sheet flanked by four helices. In both the $P2_1$ and the $C2$ crystal forms, the quaternary structure of PurS is a tetramer. The concave faces of the PurS dimers interact via the C-terminal region to form a twelve-stranded β-barrel with a hydrophilic core. We used the structure of PurS together with the structure of lgPurL from *Salmonella typhimurium* to construct a model of the PurS/smPurL/PurQ complex. The HisH (glutaminase) domain of imidazole glycerol phosphate synthetase was used as an additional model of PurQ. The model shows stoichiometry of 2PurS/smPurL/PurQ using a PurS dimer or 4PurS/2smPurL/2PurQ using a PurS tetramer. Both models place key conserved residues at the ATP/FGAR binding site and at a structural ADP binding site. The homology model is consistent with biochemical studies on the reconstituted complex.

Our laboratories have had a long-standing interest in the formylglycinamide ribonucleotide amidotransferase (FGAR-AT)[1] enzymes that catalyze the fourth step in the purine biosynthetic pathway. These enzymes catalyze the conversion of formylglycinamide ribonucleotide (FGAR), ATP, and glutamine to formylglycinamidine ribonucleotide (FGAM), ADP, P_i, and glutamate (Scheme 1). PurL is the last remaining enzyme in the purine biosynthetic pathway to resist structural characterization and is of interest for three reasons. The first is that its elusive chemistry, activation of an amide for nucleophilic attack through an iminophosphate, has yet to be documented. The second is that PurL has been proposed to be a member of a new superfamily of ATP-requiring enzymes: direct evidence for this proposal has been missing. The third is that PurLs have been proposed to be the scaffold that might allow the enzymes of the purine pathway to function as a metabolon. PurLs come in two forms. In Gram-negative bacteria and most eukaryotes, PurL is a single polypeptide of 140 kDa (designated lgPurL), and in Gram-positive bacteria and archaebacteria, it is composed of three separate proteins: PurL (∼80 kDa, designated smPurL), PurQ (∼25 kDa), and PurS (∼9 kDa) (Figure 1).

Scheme 1: The Reaction Catalyzed by Formylglycinamide Ribonucleotide Amidotransferase (FGAR-AT)



Our initial interests focused on crystallization of the more intensively investigated lgPurLs. The fruit of these efforts, the first structure of a lgPurL, is reported in an accompanying paper (*1*). The difficulties associated with obtaining crystals of a lgPurL caused us to investigate in parallel the smPurLs and their protein cohorts, PurQ and PurS. These efforts have recently resulted in successful reconstitution of the active

[†] This work was supported by NIH Grants RR-15301 (S.E.E.) and GM32191 (J.S.). SEE is indebted to the W. M. Keck Foundation and the Lucille P. Markey Charitable Trust. A.A.H. was supported by an NSF predoctoral fellowship.

[‡] The coordinates of the $C2$ and $P2_1$ crystal forms of PurS have been deposited in the Protein Data Bank under accession numbers 1T4A and ITWJ, respectively.

* To whom correspondence should be addressed. Telephone: (607) 255-7961. Fax: (607) 255-1227. E-mail: see3@cornell.edu.

[§] Cornell University.

[∥] Department of Chemistry, Massachusetts Institute of Technology.

[⊥] Departments of Chemistry and Biology, Massachusetts Institute of Technology.

[1] Abbreviations: FGAR-AT, formylglycinamide ribonucleotide amidotransferase; PEG, poly(ethylene glycol); FGAM, formylglycinamidine ribonucleotide; DEAE, diethyl aminoethyl; SeMet, selenomethionine; AIR, aminoimidazole ribonucleotide; PRPP, phosphoribosyl pyrophosphate; FGAR, formylglycinamide ribonucleotide; IGP, imidazole glycerol phosphate; EDTA, ethylenediaminetetraacetic acid; NCS, noncrystallographic symmetry; ADP, adenosine 5′-diphosphate; ATP, adenosine 5′-triphosphate; Mt, *Methanobacterium thermoautotrophicum*; St, *Salmonella typhimurium*; Bs, *Bacillus subtilis*.

10.1021/bi0491292 CCC: $27.50
Published on Web 07/23/2004

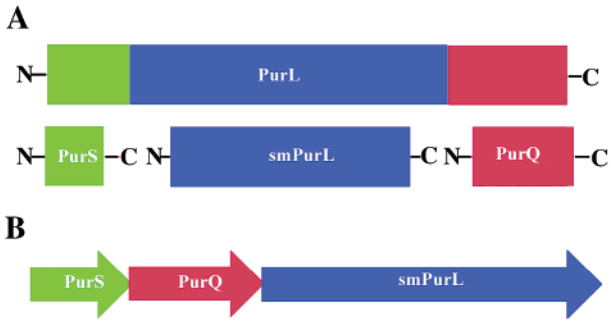


FIGURE 1: Organization of the PurSLQ protein complex and the corresponding genes. Panel A presents the alignment of lgPurL and the components of the PurSLQ complex. The N-terminal domain in lgPurL (shown in green) corresponds to the PurS dimer, the FGAM synthetase domain in lgPurL (shown in blue) corresponds to smPurL, and the C-terminal glutaminase domain in lgPurL (shown in red) corresponds to PurQ. Panel B presents the organization of the *purS*, *purQ*, and *purL* genes in the *B. subtilis* purine operon.

form of *Bacillus subtilis* FGAR-AT with the ratio of the PurS/smPurL/PurQ proteins in the complex of 2:1:1 (*2*). In the present paper, we report the structures of two different crystal forms of the *B. subtilis* small subunit, PurS, required for FGAR-AT activity.

Recently a structure of the PurS homologue from *Methanococcus thermoautotrophicum* (Mt) has resulted from a structural genomics project (*3*). The crystallographic data, supported by size exclusion chromatography (SEC) data and mutagenesis studies, suggested that the active form of MtPurS is a tetramer. Based on our recent structure of lgPurL, the biochemical studies on the corresponding *B. subtilis* smPurL, and the structural studies reported herein, we present arguments that the tetrameric form of the MtPurS cannot be the physiologicially relevant form. Our PurS structures in conjunction with the accompanying structure for the lgPurL and the glutaminase domain of imidazole glycerol phosphate synthetase have allowed us to model the PurSLQ FGAR-AT. The model suggests that the active site and the inactive ADP binding site found in lgPurL will also be present in this complex.

EXPERIMENTAL PROCEDURES

Expression and Purification of Native PurS and SeMet-PurS. PurS was purified as described previously (*2*). The purified protein was dialyzed against Millipore filtered water containing 0.02% w/v NaN_3 prior to crystallization trials. Production and purification of selenomethionine (SeMet) incorporated PurS followed the same protocol as the wild-type PurS (*2*) with a few modifications. pET-PurS was transformed into *E. coli* B834(DE3) cells (Novagen Inc.), which are auxotrophic for methionine. The cells were grown in 1 L cultures containing M9 media supplemented with 40 μg/mL L-amino acids (excluding methionine), 1× BME vitamin solution (GibcoBRL), 0.4% (w/v) glucose, 2 mM $MgSO_4$, 25 μg/mL $FeSO_4 \cdot 7H_2O$, 100 μg/mL ampicillin, 0.1 mM $CaCl_2$, and 40 μg/mL L-SeMet. Each liter of growth media was inoculated with 5 mL of a saturated starter culture containing the growth medium with L-methionine in place of L-SeMet. Before inoculation, the 5 mL of starter culture cells was pelleted and washed in the growth media to remove all traces of L-methionine. Typically 3.5 g of cells were obtained from 1 L of culture. The cells were either frozen in liquid nitrogen or used directly in purification. Cells (3.5 g) were resuspended in 20 mL of buffer and lysed in wash buffer. The supernatant was loaded onto a MonoQ HR 10/10 Pharmacia column, and the protein was eluted using a salt gradient as decribed for the wild-type PurS (*2*). The eluted protein was then loaded onto a G75 Pharmacia gel filtration column previously equilibrated with the wash buffer. The protein fractions were collected, and the protein was then dialyzed against Millipore filtered water containing 0.5 mM DTT. Typically, 50 mg of SeMet-PurS were obtained from 3.5 g of cells.

Crystallization of PurS. Initial crystallization conditions for both native and SeMet-PurS were determined using sparse matrix screens Crystal Screen 1 and Crystal Screen 2 (Hampton Research) and 20 mg/mL PurS. Native PurS was crystallized from 2.0 M ammonium sulfate, 6% PEG 400, 100 mM HEPES, pH 7.3. The crystals were grown at room temperature using the hanging-drop vapor diffusion technique. Drops (6 μL) containing a 2:1 mixture of protein and reservoir solution were optimal for crystal growth. Crystals grew over a period of 4−7 days. Under these conditions, PurS crystallizes in the monoclinic space group $P2_1$ with unit cell dimensions $a = 42.91$ Å, $b = 87.96$ Å, $c = 52.68$ Å, and $\beta = 94.97°$.

The SeMet-PurS crystals grew in 20% PEG 4K, 5% glycerol, 15% 2-propanol, 100 mM sodium citrate, pH 5.6. Optimal drop size was 4 μL using a 1:1 mixture of protein and reservoir. Under these conditions, the protein crystallizes in the space group $C2$ with unit cell dimensions $a = 89.13$ Å, $b = 42.21$ Å, $c = 47.03$ Å, and $\beta = 118.22°$.

Data Collection and Processing. The native data were collected on a flash frozen crystal at 100 K. The crystals were incubated in 0.085 M HEPES, pH 7.5, 1.7% PEG400, 1.7 M ammonium sulfate, and 15% glycerol for 2−5 s before being frozen in liquid nitrogen. X-ray intensities were measured using a MAR 345 mm image plate detector and Cu Kα X-rays from a Rigaku RU-300 rotating anode generator. The data were measured in 1° oscillation steps with 2 min exposure times and a crystal to detector distance of 200 mm for a total of 180°.

The SeMet-PurS crystal was taken directly from the drop and flash frozen in a nitrogen stream at 100 K. The data on this crystal were measured at the NE-CAT beamline 8-BM at the Advanced Photon Source (APS) using a Quantum-315 detector (Area Detector Systems Corporation) in binned format. An X-ray absorption spectrum in the vicinity of the Se K-absorption edge was determined for the SeMet-PurS crystal by recording X-ray fluorescence as a function of wavelength. Diffraction data were then collected at the peak of the spectrum (0.9791 Å). The data were measured in 1° oscillation steps with a 10 s exposure time and a crystal to detector distance of 300 mm. To minimize systematic errors, Bijvöet pairs were acquired close in time by collecting the data in 5° wedges followed by a wedge having inverse beam geometry ($\phi + 180°$). The DENZO (*4*) suite of programs was used for integration and scaling of all data. All crystals showed multiple lattices. For the final dataset, two lattices were independently indexed and integrated using space group $P1$. The datasets from the two lattices were scaled in space

Table 1: Summary of Data Collection and Processing[a]

	*C*2 form	$P2_1$ form
resolution (Å)	2.0	2.5
wavelength (Å)	0.979	1.54
no. of reflns	122 902	46 561
no. of unique reflns	10 453	13 414
redundancy	12.0	3.5
completeness	98.0 (97.3)	98.2 (99.5)
R_{sym} (%)[b]	11.0 (28.7)	9.7 (31.8)
I/σ	20.0 (10.0)	10.0 (5.0)

[a] Values for the outer resolution shell are given in parentheses. [b] $R_{sym} = \sum\sum_i|I_i - \langle I\rangle|/\sum\langle I\rangle$, where $\langle I\rangle$ is the mean intensity of the N reflections with intensities I_i and common indicies h,k,l.

Table 2: Final Refinement Statistics

	*C*2 form	$P2_1$ form
resolution (Å)	25−2.0	25−2.5
total no. of non-hydrogen atoms	1360	2597
no. of protein atoms	1276	2505
no. of water molecules	184	92
no. of reflns in refinement	19 509	13 053
no. of reflections in test set	1821	666
R factor (%)[a]	21.7	25.5
R_{free} (%)[b]	28.3	31.2
RMS deviation from ideal geometry		
bonds (Å)	0.004	0.007
angles (deg)	1.31	1.42
average B-factor (Å^2)	18.1	35.0
Ramachandran plot		
most favored region (%)	89.6	84.9
additional allowed region (%)	9.7	13.7
generously allowed region (%)	0.7	1.1
disallowed region (%)	0.0	0.4

[a] R factor $= \sum_{hkl}||F_{obsd}| - k|F_{calcd}||/\sum_{hkl}|F_{obsd}|$, where F_{obsd} and F_{calcd} are observed and calculated structure factors, respectively. [b] For R_{free}, the sum is extended over a subset of reflections (8%) excluded from all stages of refinement.

group *C*2. Final data processing statistics are shown in Table 1.

Structure Solution and Refinement. Initial phases were obtained by molecular replacement, using a polyalanine dimer model based on the MtPurS structure (PDB code 1GTD) as the search model (*3*). Rotation and translation searching were performed with the program CNS (*5*) using data to 3.0 Å resolution for the native PurS. Two dimers were found in the $P2_1$ crystal form, resulting in one PurS tetramer per asymmetric unit. The tetramer showed 222 noncrystallographic symmetry (NCS). Refinement of the molecular replacement model was carried out at 2.5 Å resolution. The initial molecular replacement solution (R factor 45%) was optimized by rigid body refinement and simulated annealing refinement. A composite omit map was then used to build side chains through clear stretches of electron density with a polyalanine model in the rest of the regions. All model building for native PurS was performed using the computer program O (*6*). Successive rounds of rigid body refinement, simulated annealing, temperature factor refinement, and model rebuilding further improved the model. The NCS constraints were kept on for initial rounds of refinement and then slowly relaxed. Water molecules were included in subsequent rounds of refinement based on the criteria that the peak in difference electron density maps was greater than 3σ and the water molecule formed at least one hydrogen bond with a protein, ligand, or water molecule. The final R factor of 25.5% (31.2% R_{free}) was high because of overlap of reflections from the two separate lattices.

The full atom native *B. subtilis* PurS structure was used as the molecular replacement search model for the *C*2 crystal form of the SeMet-PurS. Molecular replacement was performed using CNS (*5*) with the data to 3.0 Å resolution. The *C*2 form of PurS contains a dimer in its asymmetric unit. A composite omit map showed some errors in the side chain placement. After manual adjustment of the model using O, the model was further refined by the procedure described above. The final R factor after the addition of water molecules was 21.7% (28.2% R_{free}). The final refinement statistics are presented in Table 2. The Se atom positions were confirmed using an anomalous difference Fourier map but were not directly used for phase calculation.

Construction of the PurS/smPurL/PurQ Complex Model. Models of *B. subtilis* smPurL and PurQ were constructed. The BsPurL homology model was based on the FGAM synthetase domain of *S. typhimurium* lgPurL (residues 215−977; 21% identity). The PurQ homology model was based on both the glutaminase domain of lgPurL (28% identity) (*1*), which is functionally most similar to BsPurQ, and the *Thermus thermophilus* HisH (13% identity) (*7*), PDB code 1KA9, which gave the highest score other than the StPurL glutaminase domain in a PSI-BLAST (*8*) search using BsPurQ as the target sequence. The sequences were aligned using ClustalX (*9*) and the alignment was manually adjusted with a bias toward moving insertions and deletions to loop regions of the lgPurL model. Homology models were generated for BsPurQ and BsPurL using Modeller version 6 (*10*, *11*) using three cycles of slow MD annealing, corresponding to Modeller parameters MD_LEVEL = refine_3, LIBRARY_SCHEDULE = 2. Prior to molecular mechanics minimization, the PurQ and smPurL models were perturbed by applying random shifts between −4 and 4 Å to the non-hydrogen atoms. Ten models were generated in this way for each PurQ and smPurL. Hydrogen atoms were added with ideal geometry and prior to full atom minimization the hydrogen atoms of freely rotating groups were adjusted by molecular mechanics energy minimization to remove bad contacts. After full atom energy minimization, the lowest energy model for each PurQ and smPurL was chosen for modeling the PurSLQ complex. The rms deviations between the lowest energy model and the other nine models ranged from 1 to 1.5 Å.

The structure of lgPurL from *S. typhimurium* was used as a template for assembling the PurS, smPurL, and PurQ components. The lowest-energy homology models and a dimer from the *B. subtilis* PurS crystal structure were superimposed onto the lgPurL using matrices calculated by DALI (*12*). Unfavorable steric contacts were resolved by individually minimizing the interface regions between each protein pair while freezing all other atoms. The interface regions were defined as all residues having one atom (including hydrogen atoms) within 3.0 Å of any atom in the other protein. A final minimization of the entire structure was performed with 10 kcal/(mol·Å^2) restraints on all non-hydrogen atoms.

A 4PurS/2smPurL/2PurQ complex was generated by superimposing the two 2PurS/smPurL/PurQ complexes onto the crystallographic tetramer of PurS from the $P2_1$ crystal

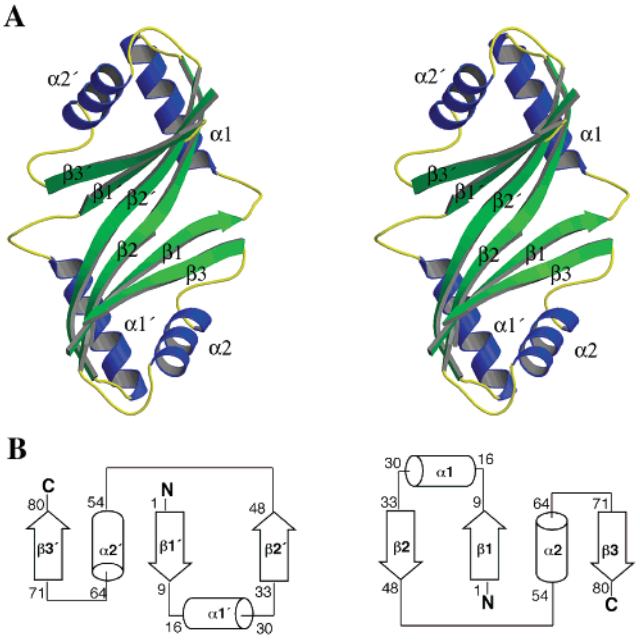


FIGURE 2: Structure of *B. subtilis* PurS: (A) stereodiagram of the PurS dimer with α-helices in blue and β-strands in green; (B) topology diagram for the PurS dimer. The secondary structural elements are labeled and the numbers for the first and last residues are given. The figure was prepared using MOLSCRIPT (*16*−*18*) and RASTER3D (*19*).

form. The interface region between the two PurS dimers was then minimized. The point group symmetry of the PurS tetramer allows for two possible orientations of the second 2PurS/smPurL/PurQ complex. Both orientations were examined and neither resulted in unfavorable interactions between the two halves. Energy minimization of the complex was carried out using the GB/SA solvation model (*13*) and AMBER94 force field (*14*) implemented in version 7.2 of the program MacroModel (*15*). Structures were minimized to a gradient of 0.05 kJ/(mol·Å).

RESULTS

PurS Dimer. The three-dimensional structure of PurS was determined by molecular replacement using a MtPurS dimer as the search model (*3*). *B. subtilis* PurS crystallizes in two crystal forms, $P2_1$ and $C2$, under different crystallization conditions: one at pH 5.6 and the other at 7.3. The PurS monomer consists of 84 amino acids of which the first 80 are visible in the experimental electron density maps for both the $P2_1$ and $C2$ crystal forms. In the $C2$ crystal form, there are two molecules in the asymmetric unit, whereas in the $P2_1$ crystal form, there are four molecules in the asymmetric unit. The PurS monomer consists of three antiparallel β-strands flanked by two helices (Figure 2). Helix α1 extends away from the β-sheet structure and is involved in interactions with the adjacent PurS monomer, whereas helix α2 packs against strand β3.

PurS forms a tight dimer (2800 Å^2 buried surface area), which is generated by the noncrystallographic 2-fold axes in both the $C2$ and $P2_1$ space groups. The dimer interface is formed primarily by extensive hydrogen bonding interactions between strands β2 and β2′ (residues A33−A48 and B33−B48). These strands extend through the entire width of two monomers and form 12 hydrogen bonds. These strands also are part of the three-stranded β-sheet found in each monomer. All four helices are located on the convex face of the dimer. They form a claw-like structure with a depression in the center. The dimers from each space group are almost identical.

PurS Tetramer. In both the $P2_1$ and the $C2$ crystal forms, the quaternary structure of PurS is a tetramer. In the case of the $P2_1$ crystal form, the tetramer is formed by two PurS dimers (A/B and C/D) that are related by a NCS 2-fold axis (Figure 3A). However, in the case of the $C2$ crystal form, the tetramer is formed by two dimers (AB and A′B′) that are related by a crystallographic 2-fold axis (Figure 3B). For both tetramers, the concave faces of the PurS dimers interact via the C-terminal regions (strand β3) to form a twelve-stranded β-barrel. The core of the β-barrel is very hydrophilic and is lined with residues Glu77, Arg75, Lys5, Tyr7, and Tyr42 from each of the four monomers. The helices are located on the convex surface of the tetramer.

The hydrogen-bonding interactions responsible for the formation of the tetramer are different in the two crystal forms (Figure 4A,B). In the case of the $P2_1$ crystal form (Figure 4A), the tetramer is stabilized by four hydrogen bonds formed between backbone atoms. The carbonyl oxygen and the amide nitrogen atoms of Val78A and Val78D are hydrogen bonded to the amide nitrogen and oxygen atoms of Tyr76D and Tyr76A, respectively.

The halves of the tetramer in the $C2$ crystal form are rotated by approximately 10° with respect to the tetramer of the $P2_1$ crystal form (Figure 4B). This results in a shift of the hydrogen bonding pattern by two residues with respect to the $P2_1$ form. The tetramer is stabilized by two hydrogen bonding interactions between the backbone carbonyl oxygen atoms of Val78A and the amide nitrogen atoms of Val78D from the symmetry-related subunit. The tetramer is further stabilized by bridging water molecules that form hydrogen bonds with the carbonyl oxygen atom of Tyr76 and the amide nitrogen atom of Glu80.

Crystal Packing Contacts. The two crystal forms of PurS show very different crystal packing interactions. The PurS tetramers are more densely packed in the $C2$ crystal form (solvent content 47%) than $P2_1$ crystal form (solvent content 60%). The tetramer of the $P2_1$ crystal form packs against an adjacent tetramer by interactions between helix α1 and the loop 49−53. This is in contrast to the MtPurS structure in which α1 is involved in tetramer formation (*3*) (Figure 3C). The backbone carbonyl atom of Leu15A in helix α1 hydrogen bonds to the NE1 nitrogen atom of His24C from the adjacent tetramer. Hydrogen bonds also form between the backbone amide nitrogen atom of Asp53D and the OE1 oxygen of Asp51D.

The tetramer in space group $C2$ makes crystal packing contacts with 10 different tetramers. The crystal contacts primarily involve residues in helix α1 and helix α2. The crystal packing interface is mostly hydrophilic and includes several ordered water molecules. Key hydrogen bonding interactions include the hydroxyl group of Ser20B to the OE2 oxygen atom of Glu60A and NZ nitrogen atom of Lys64A to the amide oxygen atom of Gln23B. There are also hydrogen bonds between Asn32B and Thr69A, His32B, and Thr69A, Arg49B with Asp53A, Asp53B with Arg49A, Glu55B with Tyr2A, and Glu76B with Glu80A.

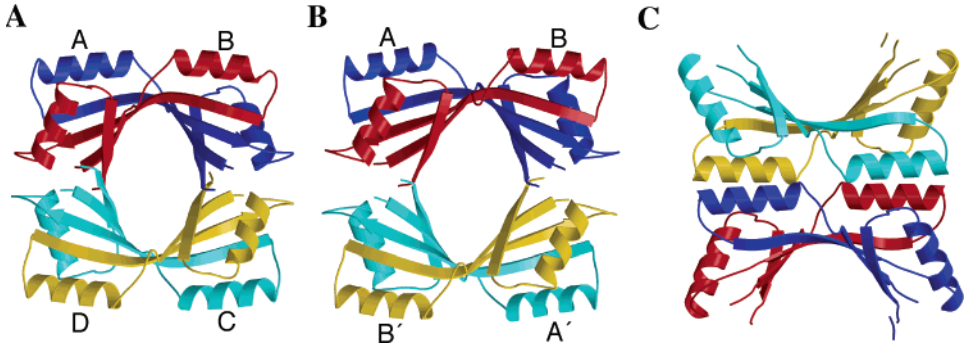


FIGURE 3: Quaternary structure of *B. subtilis* PurS: (A) the tetrameric structure of *B. subtilis* PurS as formed in space group $P2_1$; (B) the tetrameric structure of *B. subtilis* PurS as seen in space group *C*2. Each tetramer has 222 point symmetry with the 2-fold axes aligned vertically, horizontally, and perpendicular to the page. The red/blue pair and the yellow/cyan pair form the dimers shown in Figure 2. The hydrogen bonding between red and cyan monomers and the blue and yellow monomers are different for the two tetramers (see Figure 4 and the text for a description). Panel C presents the tetrameric structure reported for MtPurS (*3*). The figure was prepared using MOLSCRIPT (*16*−*18*) and RASTER3D (*19*).

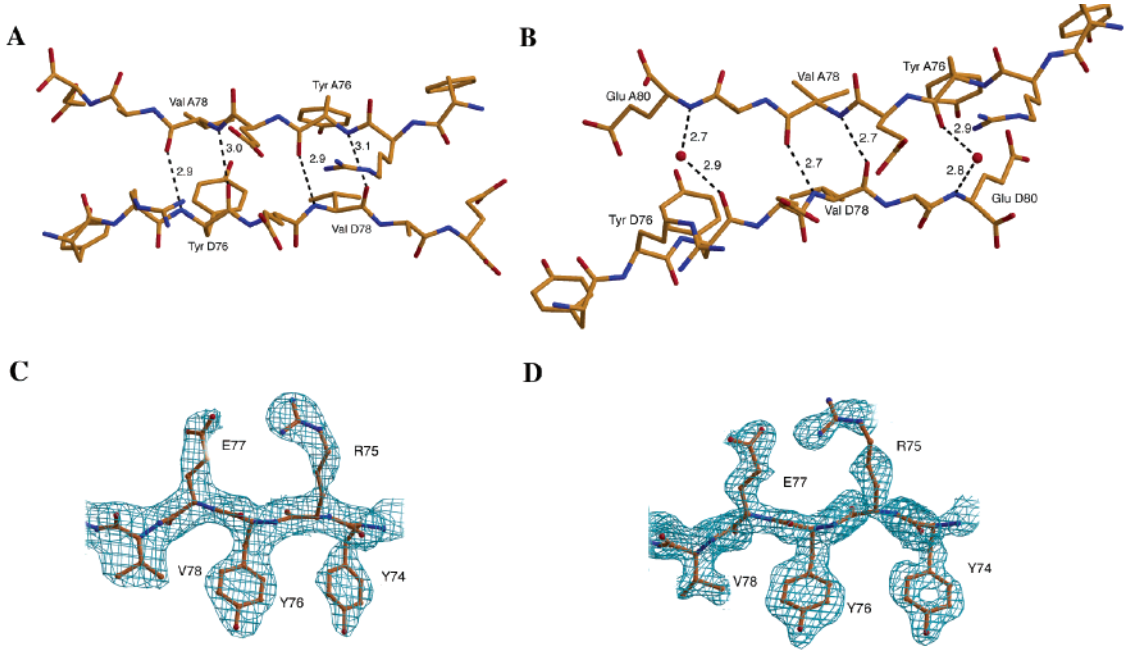


FIGURE 4: Hydrogen bonding scheme at the tetramer interface ($\beta3/\beta3'$). Panel A presents hydrogen bonds for the tetramer in space group $P2_1$. Panel B presents the hydrogen bonds for the tetramer in space group *C*2. The monomer designations correspond to the labeling in Figure 3. Panel C shows a section of the $2F_o - F_c$ electron density in the vicinity of the tetramer interface in space group $P2_1$ contoured at 1.5 σ. Panel D shows a section of the $2F_o - F_c$ electron density in the vicinity of the tetramer interface in space group *C*2 contoured at 1.5 σ. The figure was prepared using MOLSCRIPT (*16*−*18*) and RASTER3D (*19*).

Model of PurSQL Complex. Our recent biochemical studies on smPurLs have demonstrated that an active complex is formed from PurS/smPurL/PurQ in a ratio of 2:1:1. Our SEC studies did not, however, allow a definitive distinction between this stoichiometry and 4:2:2 stoichiometry (*2*). The observation of two tetrameric BsPurSs (Figure 3A,B) would seem to support the latter option. The biochemical data, our recent structure of lgPurL, the wealth of knowledge about the class I glutaminase domains, and the *B. subtilis* structures of PurS offer the opportunity to model the smPurL FGAR-AT.

Initially we made a homology model of the PurQ and smPurL proteins. For PurQ, the *Thermus thermophilus* HisH, the glutaminase domain of imidazole glycerol phosphate synthase, which has 13% sequence identity with PurQ, and the C-terminal glutaminase domain of lgPurL, which shows 28% sequence identity with PurQ, were employed. Structure-based sequence alignments and the structures of HisH and the StPurL glutaminase domain (residues 1018−1294) were used with Modeller (*11*) to generate the PurQ model shown in Figure 5A (red). More challenging was the model of the smPurL protein (21% identity), which contains many more loops than PurQ. The structure of this domain within the lgPurL identified four subdomains (A1, B1, A2, and B2). Structure-based sequence alignment of all smPurLs with the corresponding domain (residues 215−977) of the lgPurL were generated (Supporting Information). The modeling results are shown in Figure 5A and the exploded view of the complex in Figure 5B.

For the PurS component, the BsPurS X-ray structure was used. Amazingly, the BsPurS showed no significant sequence similarity to any part of StPurL. There is however, a striking structural similarity between the N-terminal domain of StPurL and the BsPurS dimer (Figure 5C). The basic topology is a half-barrel maintained by extensive hydrogen bonding interactions between strands $\beta2$ and $\beta2'$ in the case

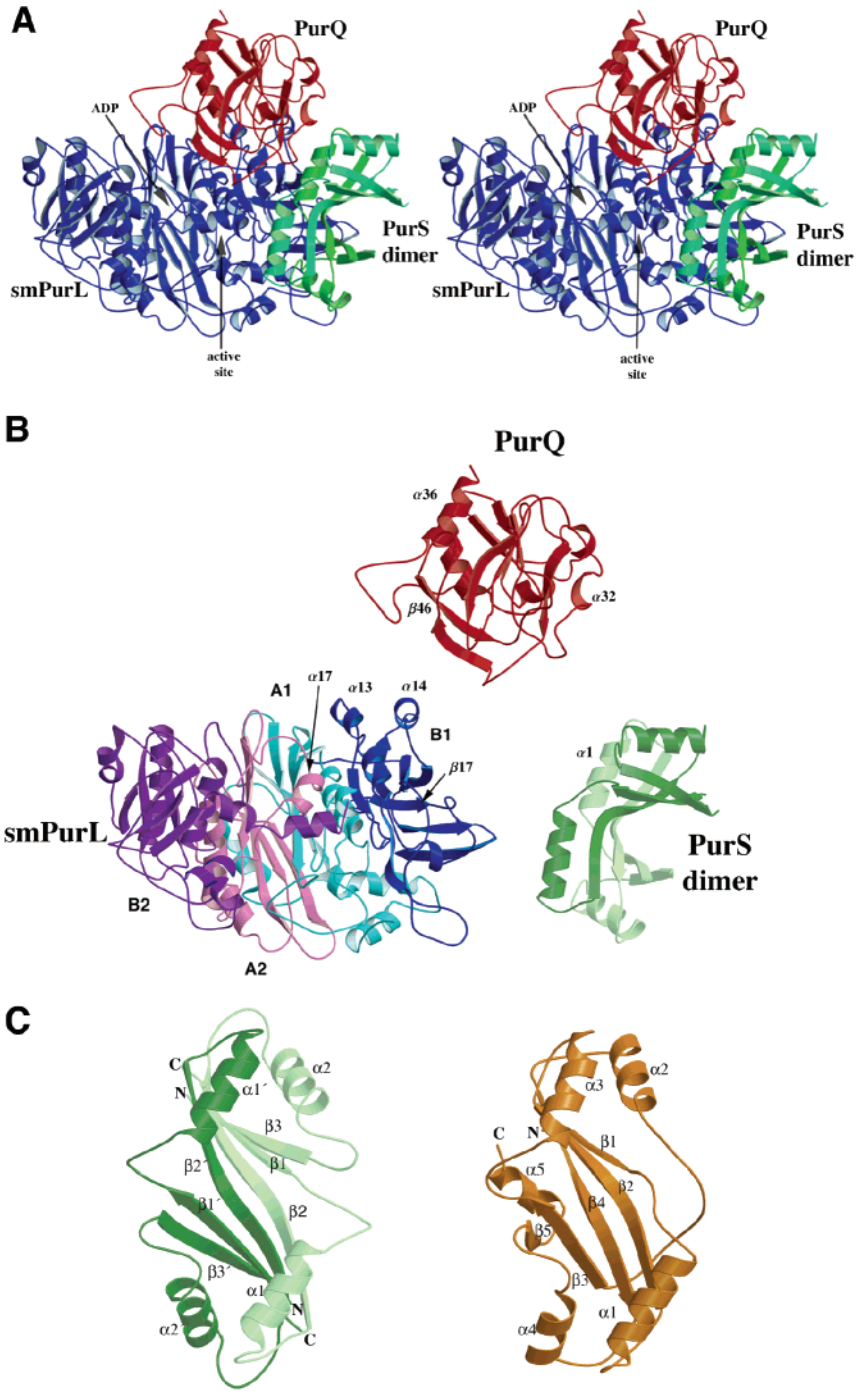


FIGURE 5: Predicted structure of the PurSLQ complex. Panel A presents a stereoview of the model of the 2PurS/smPurL/PurQ complex. The model consists of a PurS dimer shown in light green/dark green, smPurL shown in blue, and PurQ shown in red. The positions of the PurQ and smPurL actives site are labeled. The location of the structural ADP molecule is also indicated. Panel B presents an exploded view the PurSLQ complex. Subdomains of smPurL are indicated (A1 = cyan, B1 = blue, A2 = pink, and B2 = magenta). Structural elements involved in protein/protein interactions are labeled using the secondary structure designations for StPurL in the case of the smPurL and PurQ domains (see Supplementary Figure 1). Panel C presents a comparison of the PurS dimer and the N-terminal portion (residues 1−140) of StPurL. The view is normal to the 2-fold axis of the PurS dimer and the pseudo-2-fold axis of the StPurL N-terminal domain. Secondary structural elements and the N- and C-termini are labeled. The figure was prepared using MOLSCRIPT (*16*−*18*) and RASTER3D (*19*).

of the BsPurS dimer and strands β2 and β4 in the case of StPurL. In both folds, the central β-sheet core is flanked by four helices that are involved in interactions with neighboring domains in the case of StPurL and PurQ and smPurL in the case of BsPurS. One difference is that the PurS dimer has additional edge strands, β3 and β3′, and forms a six-stranded half-barrel, while the N-terminal domain of StPurL forms a five-stranded half-barrel. In the case of PurS, the additional β-strands allow two half-barrels to join into a tetrameric structure shown in Figure 3.

The initial model of a 2PurS/smPurL/PurQ complex (Figure 5A) was constructed by superimposition of the *B. subtilis* PurS dimer and the smPurL and PurQ homology models onto the crystal structure of StPurL (*1*). Multiple homology models were generated by randomization for the PurQ and smPurL proteins. The models all showed the same

general features after minimization, and the lowest energy models of PurQ and smPurL were chosen for assembly of the PurSLQ complex. The individual proteins superimposed easily on the StPurL template resulting in very few bad contacts between protein modules. The final model of the PurSLQ complex shows a slight gap of a few angstroms between PurS and PurQ proteins, corresponding to a gap between the N-terminal and glutaminase domains of StPurL. Closing of the gap is possibly associated with ammonia channel formation in the active form of the complex. Although the gap could be closed by manual adjustment, it is retained in the model that we describe here.

Formation of the 2:1:1 complex buries 1437 Å^2 at the PurS/PurQ interface, 1344 Å^2 at the PurS/smPurL interface, and 2965 Å^2 at the smPurL/PurQ interface. The binding of PurQ and smPurL to the PurS dimer is asymmetric with one of the PurS monomers (light green in Figure 5) providing most of the interactions with PurQ and smPurL. The main interaction between the PurS dimer and PurQ is through helix α1 (light green) of PurS, which spans residues 16−30, with the loop preceding helix α32 in PurQ (PurQ and smPurL structural descriptions correspond to StPurL numbering as shown in Supplementary Figure 1). Residues 10−15, preceding helix α1 in PurS, interact with smPurL via a stretch of 10 residues following *β*17 in the B1 domain. A loop preceding α17 in the A2 domain of smPurL interacts with a loop following α36 in PurQ. The N-terminal portion of PurQ also interacts with residues from α20 in the B2 domain of smPurL. Helix α13 in the B1 domain interacts with residues following *β*26 in PurQ, and helix α14 interacts with residues preceding α32. The active sites of smPurL and PurQ are separated by about 30 Å.

4PurS/2smPurL/2PurQ Complex Model. The observation of a tetrameric structure in two different crystal forms of BsPurS suggested the possibility of a 4PurS/2smPurL/2PurQ complex. We constructed models for this complex by superimposing two 2PurS/1smPurL/1PurQ complexes on the BsPurS tetramer from the $P2_1$ crystal form. Two arrangements are possible: a cis complex (Figure 6A) in which the PurQ domains are on the same sides of the barrel formed by the tetrameric core and a trans complex (Figure 6B) in which the PurQ domains are on opposite sides of the tetrameric core. No unfavorable interdomain contacts are created in the assembly of either the cis or the trans 4PurS/2smPurL/2PurQ complex, and no interference with ligand binding sites is observed.

DISCUSSION

Comparison of Dimeric and Tetrameric Structures of BsPurS and MtPurS. Despite the low sequence identity between MtPurS and BsPurS (25%), the structures of BsPurS and MtPurS both form tight dimers with a central six-stranded *β*-sheet flanked by four helices. The dimer interface is formed by interactions between the central *β*-strands, *β*2 and *β*2′, each 12 residues long and related by a 2-fold axis. Two monomers form an interdigitated dimer in which helix α1 and the first half of strand *β*2 insert into the adjacent monomer (Figure 2A). The concave surface of the six-stranded *β*-sheet is exposed, and the convex surface is flanked by the four α-helices. A structural superposition of the MtPurS dimer onto the BsPurS dimer using DALI (*12*)

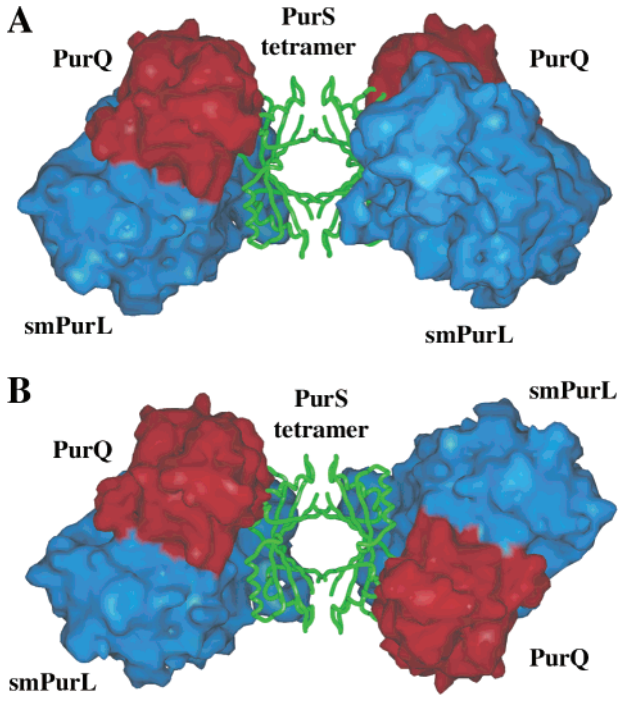


FIGURE 6: Model of the 4PurS/2smPurL/2PurQ complex: (A) the cis complex; (B) the trans complex. The view is along the axis of the barrel formed by the PurS tetramer. The PurS tetramer is shown in green as a tube representation. PurQ and smPurL are shown as surface diagrams in red and blue, respectively. The figure was prepared using MOLSCRIPT (*16*−*18*) and RASTER3D (*19*).

with 158 Cα positions gives an rms deviation of 2.7 Å for the $P2_1$ form and 2.8 Å for the *C*2 crystal form. Between the two structures, the helices align better than the central *β*-core, and the major structural differences are in a loop spanning residues 49−53, which contains a four residue gap in BsPurS relative to MtPurS. This loop also shows significant variation in the two crystal structures of BsPurS. The differences may result from crystal packing or from differences in the pH of crystallization (pH 7.3 for the $P2_1$ form and pH 5.6 for the *C*2 form). Sequence alignment of 14 different PurS sequences also shows high sequence variation in this loop.

Despite the similarities in the dimeric structures of the BsPurS and MtPurS, the tetrameric structures are very different (Figure 3A,B compared to Figure 3C). In the MtPurS structure, the tetramer is stabilized by hydrophobic interactions of the two pairs of helices through the convex surface of the dimer (Leu11, Leu16, Ala21, Ile23, Leu29, and Leu 30 from helix α1 and Leu66, Leu67, Leu71, and Leu72 from helix α2). Analysis of the tetramers for the $P2_1$ and *C*2 crystal forms of BsPurS shows no similar interactions to those found in MtPurS. In fact one of the four α1 helices required for tetramer formation in the MtPurS (Figure 3C) is involved in direct interaction of PurS with smPurL in our modeled complex structure (Figure 5A). Thus the tetrameric form of MtPurS is unlikely to be a part of the physiologically active form of FGAR-AT. It is possible that the presence of a His-tag in MtPurS disrupts the biologically relevant tetramer by inserting between the dimer halves (*3*). However, this possible artifact does not exist for BsPurS, which was expressed in a non-His-tag construct.

Evidence for a Single Active Site in smPurLs. Our structure of lgPurL provided strong evidence for a gene duplication

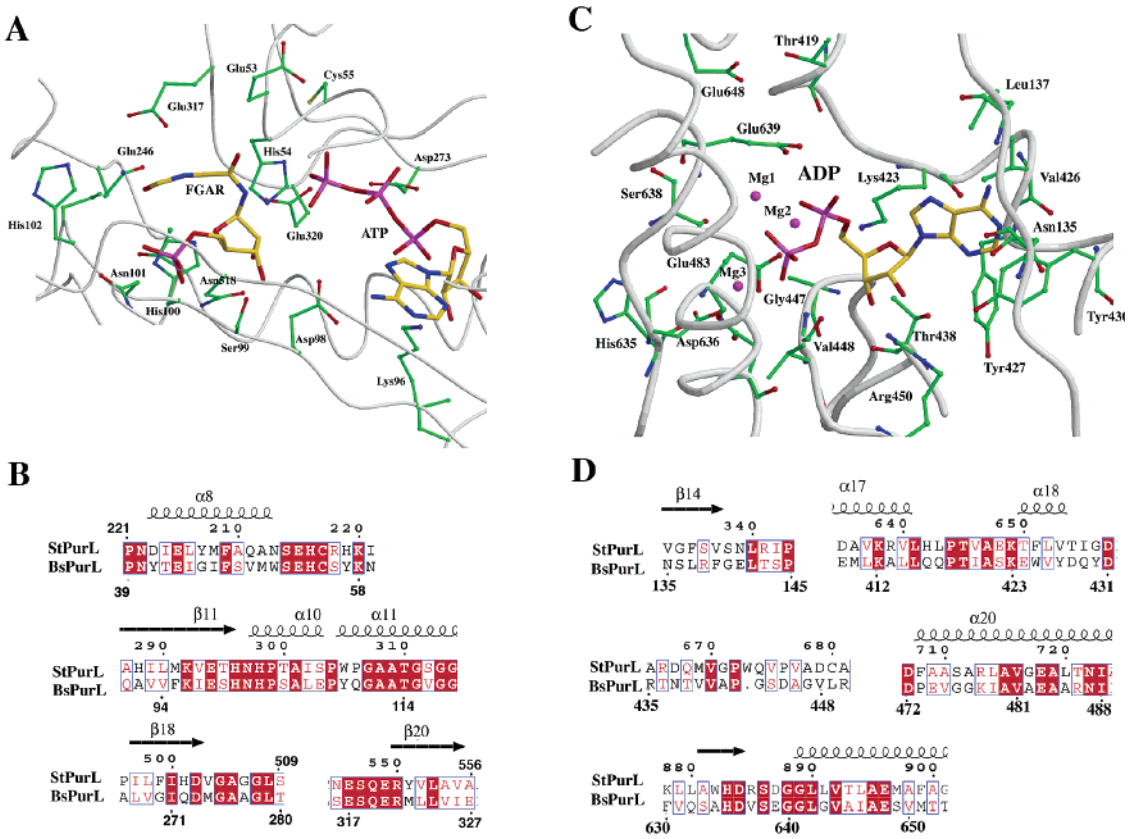


FIGURE 7: Ligand binding sites for the PurSLQ complex. FGAR, ATP, and ADP were generated by superposition of the ligands from StPurL using FGAR and ATP molecules that were modeled into the lgPurL active site (*1*). Panel A shows the FGAR and ATP binding site. Panel B shows selected regions of sequence alignment between BsPurL and StPurL, corresponding to the FGAR/ATP binding site. Panel C shows the MgADP binding site. Panel D shows selected regions of sequence alignment between BsPurL and StPurL, corresponding to the structural ADP binding site. Conserved residues within a 5 Å radius of the ligands are shown in a ball-and-stick representation in panels A and C. The figure was prepared using MOLSCRIPT (*16*−*18*) and RASTER3D (*19*).

event in the FGAM synthetase domain that by itself would suggest the possibility of two active sites. However, the structure further revealed an unexpected tightly bound ADP in one of these sites, suggesting that only the second site was involved in catalysis. Structure-based sequence alignments required for the homology model of the smPurL (Figure 7A,B, and Supplementary Figure 1) provide strong support for identification of only a single active site where FGAR and ATP bind (Figure 7A). The homology model of BsPurL preserves the key structural features of the central FGAM synthetase domain of StPurL where the active site lies between the A1 and B1 subdomains (*1*). The alignment of 26 smPurL sequences reveals that the N-terminal half (A1 and B1 domains, Figure 5B) contains 63 fully conserved residues out of 350 total. In the other putative active site of the homology model, between the A2 and B2 subdomains (Figure 5B), the number of fully conserved residues is only 17. The higher degree of conservation in the A1 and B1 subdomains is consistent with an FGAM synthetase active site residing in the N-terminal half of smPurL.

Putative active site residues that are absolutely conserved among both smPurLs and lgPurLs are found in four segments of the polypeptide chain. Ser214, Glu215, and His216 in StPurL correspond to residues Ser52, Glu53, and His54 in BsPurL. These residues are on a loop thought to cap the putative binding site for FGAR. Residues 294−297 in StPurL structurally align with residues 98−102 in BsPurL and are proposed to line the entrance to the active site cleft and bind the 5′-phosphate of FGAR. Residues 447−450, a disordered, glycine-rich region thought to be a P loop for phosphate binding in StPurL, align with residues 115−118 in BsPurL. This region is proposed to assist ATP binding. Residues 502−507 in StPurL, corresponding to residues 273−277 in BsPurL, contain a conserved aspartate residue, which is possibly involved in metal binding. Residues Lys292, Pro299, Gln471, Glu546, and Glu549 in StPurL also align with residues Lys96, Pro103, Gln245, Glu317, and Glu320 in BsPurL. Active site residues 245−250 and residues 316−322 are completely conserved among all smPurLs but different in lgPurL.

A superposition of the PurSLQ complex onto the crystal structure of StPurL reveals a possible channel similar to the one formed by the N-terminal, A1, and B1 domains of StPurL. The analogous channels in BsPurSLQ and StPurL have similar residues near the entrance of the FGAR/ATP binding site; residues Arg121, Ile131, and Glu252 in BsPurL correspond to Arg317, Ile316, and Glu478 in StPurL. In addition, Phe1094 in StPurL, which probably acts as a gate to the glutaminase domain in StPurL, corresponds to residue Phe52 in BsPurQ. The interiors of the channels show considerable variation, both between large and small PurLs and within the small PurLs.

The PurSLQ model provides an explanation for the loss of one of the smPurL active sites following gene duplication. As in the case of lgPurL, the homology model of smPurL from *B. subtilis* shows 2-fold pseudosymmetry suggesting gene duplication (*1*). The PurSLQ complex model as with the sequence alignments suggests that only one of the halves

is catalytically active. There is only one glutaminase domain, and it contacts the N-terminal half of smPurL. In addition, the entrance to the possible active site cleft in the C-terminal half is blocked by the helix connecting the A2 and the B2 domains. Using the 2-fold pseudosymmetry to model a PurQ and a PurS on the A2/B2 sides results in blocking the FGAR entry point into the active site on the A1/B1 side of smPurL. Therefore, if smPurL were to contain two active sites, the PurQ domain of the 2-fold related site would block entry of FGAR. Utilization of both potential active sites would require both PurQ and smPurL proteins to bind their substrates before complex formation and complete dissociation of the complex to release products.

Evidence of an ADP Binding Site in smPurL. The crystal structure of lgPurL shows a deeply buried ADP molecule plus three Mg^{2+} ions bound in the C-terminal half of the FGAM synthetase domain (*1*). The ADP binding site is related to the FGAM synthetase active site by 2-fold pseudosymmetry. Despite the low sequence conservation in the C-terminal half of smPurLs with the lgPurLs noted above, there are sufficient similarities to generate a proposed ADP binding site for PurSLQ (Figure 7C,D). The conserved residues Glu483, Asn487, Asp636, Glu639, and Glu648 in the vicinity of the Mg^{2+} binding site in BsPurL, for example, align with conserved Mg^{2+} binding residues Glu718, Asn722, Asp884, Asp887, and Glu896 in StPurL. Most of the strong interactions occur between the α- and β-phosphates of ADP and the Mg^{2+} ions, which in turn are stabilized by hydrogen bonding interactions with the protein either directly or through water molecules.

Similar to StPurL, the adenine base of ADP is located in a hydrophobic pocket formed in BsPurL by residues Leu137, Val426, Ty427, Tyr430, and Val334. In StPurL, the adenine is further stabilized by hydrogen bonding interactions of the N3 nitrogen atom with Asn668. In the case of BsPurL, the N1 nitrogen atom of the adenine base hydrogen bonds to Asn135. Sequence comparison of smPurLs shows that this residue is always aspartate or asparagine. Conserved Lys649, which helps to stabilize the α-phosphate of ADP, is located at the C-terminus of the helix that bridges the two halves of the StPurL FGAM synthetase domain. This residue is Lys423 in BsPurL and is also conserved among smPurLs. Among the small and large PurLs, there is some variability in the ribose binding region. In BsPurL, a conserved Lys450 is positioned to bind the ADP sugar moiety, but there is no counterpart of this residue in StPurL. StPurL utilizes the backbone nitrogen atom of Ala678 to hydrogen bond to the O3′ hydroxyl of the ribose group of ADP, and Gly447 in BsPurL is positioned to form an analogous interaction.

In the region where Mg^{2+}/ADP binds in StPurL, core β-strand 13 does not align structurally with BsPurL and is much shorter in length. Residues 377−409 that form a loop-strand-loop moiety in StPurL have no structural counterpart in smPurL. The sequence alignment of lgPurL with that of smPurL reveals a 15-residue gap in this region. A six-residue strand in smPurL replaces the loop-strand-loop moiety present in this region. As a result, the central β-barrel in the case of smPurL is wider than that of lgPurL. The observation of a structural ADP molecule in the crystal structure of StPurL (*1*), and the location of conserved residues in the PurSLQ model suggest a similar ADP binding site for smPurL. Recent biochemical studies have now shown that ADP is required for formation of the *B. subtilis* PurSLQ complex (*2*).

4PurS/2smPurL/2PurQ vs 2PurS/1smPurL/1PurQ Complex? The two crystal forms of *B. subtilis* PurS show two different but closely related tetramers. In addition, the MtPurS can also form yet another tetrameric structure. The different crystal forms demonstrate the plasticity of the tetrameric interface of which the buried surface area is about 50% that of the dimeric interface. Size exclusion chromatography experiments on the MtPurS have been reported to support a tetrameric structure (*3*), while SEC on BsPurS gave an ambiguous result (30 kDa for a monomer of 10 kDa) (*2*). SEC of the *B. subtilis* FGAR-AT also gave ambiguous results (153 kDa, while a 2:1:1 complex would be 125 kDa). The crystallographic data reported herein and our modeling studies show that a 4PurS/2smPurL/2PurQ complex can be constructed using the tetrameric PurS without creating unfavorable interaction components. Thus while the crystallographic data support a 4:2:2 structure, further biophysical methods are required to distinguish between this structure and the 2:1:1 complex in solution under physiological conditions.

ACKNOWLEDGMENT

We thank Ms. Leslie Kinsland for assistance in the preparation of this manuscript. We thank the NE-CAT 8BM beam line supported by NIH Grant RR15301 for provision of beam time. We thank Dr. Cynthia Kinsland of the Cornell Protein Production Facility for cloning StPurL. We thank members of the Drennan laboratory at MIT for help in X-ray data collection. The Multiuser Facility for the Study of Complex Macromolecular Systems at MIT (Grants NSF-0070319 and NIH GM68762) is also gratefully acknowledged.

SUPPORTING INFORMATION AVAILABLE

Structure-based sequence alignment of all smPurLs with the corresponding domain (residues 215−977) of the lgPurL. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Anand, R., Hoskins, A. A., Stubbe, J., and Ealick, S. E. (2004) Domain Organization of *Salmonella typhimurium* Formylglycinamide Ribonucleotide Amidotransferase Revealed by X-ray Crystallography, *Biochemistry 43*, 10328−10342.
2. Hoskins, A. A., Anand, R., Ealick, S. E., and Stubbe, J. (2004) The Formylglycinamide Ribonucleotide Amidotransferase Complex from *Bacillus subtilis*: An Example of Metabolite-Mediated Complex Formation, *Biochemistry 43*, 10314−10327.
3. Batra, R., Christendat, D., Edwards, A., Arrowsmith, C., and Tong, L. (2002) Crystal structure of MTH169, a crucial component of phosphoribosylformylglycinamidine synthetase, *Proteins 49*, 285−288.
4. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307−326.
5. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D54*, 905−921.

6. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. A47*, 110−119.
7. Omi, R., Mizuguchi, H., Goto, M., Miyahara, I., Hayashi, H., Kagamiyama, H., and Hirotsu, K. (2002) Structure of imidazole glycerol phosphate synthase from *Thermus thermophilus* HB8: open-closed conformational change and ammonia tunneling, *J. Biochem (Toyko) 132*, 759−765.
8. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389−3402.
9. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res. 24*, 4876−4882.
10. Marti-Renom, M. A., Yerkovich, B., and Sali, A. (2002) Comparative protein structure prediction, in *Current Protocols in Protein Science* (Coligan, J. E., Ed.) pp 2.9.1−2.9.22, John Wiley & Sons, Inc., New York. & Sons, Incorporated.
11. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol. 234*, 779−815.
12. Holm, L., and Sander, C. (1998) Touring protein fold space with Dali/FSSP, *Nucleic Acids Res. 26*, 316−319.
13. Qiu, D., Shenkin, P. S., Hollinger, F. P., and Still, W. C. (1997) The GB/SA Continuum Model for Solvation. A Fast Analytical-Method for the Calculation of Approximate Born Radii, *J. Phys. Chem. A 101*, 3005−3014.
14. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc. 117*, 5179−5197.
15. Mohamadi, F., Richards, N. G. J., Guida, W. C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., and Still, W. C. (1990) MacroModel − an Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics, *J. Comput. Chem. 11*, 460−467.
16. Esnouf, R. (1997) An extensively modified version of Molscript which includes greatly enhanced colouring capabilities, *J. Mol. Graphics 15*, 132−134.
17. Esnouf, R. M. (1999) Further additions to MolScript version 1.4, including reading and contouring of electron-density maps, *Acta Crystallogr. D55*, 938−940.
18. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−950.
19. Merritt, E. A., and Bacon, D. J. (1997) Raster3D: Photorealistic Molecular Graphics, *Methods Enzymol. 277*, 505−524.

BI0491292