Structures of SAICAR synthetase (PurC) from *Streptococcus pneumoniae* with ADP, Mg²⁺, AIR and Asp

Nina M. Wolf,[a] Celerino Abad-Zapatero,[b] Michael E. Johnson[b] and Leslie W.-M. Fung[a]*

[a]Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA, and [b]Center for Pharmaceutical Biotechnology, University of Illinois at Chicago, Chicago, IL 60607, USA

Correspondence e-mail: lfung@uic.edu

Streptococcus pneumoniae is a multidrug-resistant pathogen that is a target of considerable interest for antibacterial drug development. One strategy for drug discovery is to inhibit an essential metabolic enzyme. The seventh step of the *de novo* purine-biosynthesis pathway converts carboxyaminoimidazole-ribonucleotide (CAIR) and L-aspartic acid (Asp) to 4-(*N*-succino)-5-aminoimidazole-4-carboxamide ribonucleotide (SAICAR) in the presence of adenosine 5′-triphosphate (ATP) using the enzyme PurC. PurC has been shown to be conditionally essential for bacterial replication. Two crystal structures of this essential enzyme from *Streptococcus pneumoniae* (*sp*PurC) in the presence of adenosine 5′-diphosphate (ADP), Mg²⁺, aminoimidazoleribonucleotide (AIR) and/or Asp have been obtained. This is the first structural study of *sp*PurC, as well as the first of PurC from any species with Asp in the active site. Based on these findings, two model structures are proposed for the active site with all of the essential ligands (ATP, Mg²⁺, Asp and CAIR) present, and a relay mechanism for the formation of the product SAICAR is suggested.

1. Introduction

Antimicrobial resistance in the treatment of pneumococcal infections has become an increased concern and is the subject of several recent reviews (Cornick & Bentley, 2012; Lynch & Zhanel, 2010; van der Poll & Opal, 2009). The bacterium *Streptococcus pneumoniae* is responsible for more than 1.6 million deaths annually (Lynch & Zhanel, 2010) from diseases such as meningitis, bacteremia, sepsis, cellulitis and pneumonia. In developed countries, severe community-acquired pneumonia remains the most common cause of death from infection, and strains of this microorganism with resistance to penicillin, β-lactams, fluoroquinolones, macrolides, vancomycin, trimethoprim and several other antimicrobials have emerged (van der Poll & Opal, 2009). The growing need for a more effective treatment of *S. pneumoniae* infection is apparent. One strategy of antibacterial development is to target essential biochemical pathways. The *de novo* purine- and pyrimidine-biosynthesis pathways have been shown to be two of the most critical pathways necessary for pathogenicity of a microorganism in blood, and the enzymes involved in these pathways have been suggested as antibiotic targets (Samant *et al.*, 2008). In 'step 7' of the *de novo* purine-biosynthesis pathway, a carbon–nitrogen bond is formed between 4-carboxy-5-aminoimidazole ribonucleotide (CAIR) (Fig. 1) and L-aspartic acid (Asp) to give 4-(*N*-succino)-5-aminoimidazole-4-carboxamide ribonucleotide (SAICAR) in the presence of adenosine 5′-triphosphate (ATP), using

the ligase enzyme phosphoribosyl-aminoimidazole-succino-carboxamide synthetase [PurC, also named 4-(*N*-succino)-5-aminoimidazole-4-carboxamide ribonucleotide synthetase; EC 6.3.2.6; Zhang, Morar *et al.*, 2008]. To rationally develop antibiotics targeting PurC, we need detailed information on the structures of PurC with its ligands and on its reaction mechanism.

We report the structures of *S. pneumoniae* PurC (*sp*PurC) with ADP, Mg^{2+}, Asp and/or AIR in its active site. The binding specifics of these molecules in the active site allow us to model the active-site structure with ligands with sufficient certainty to propose a relay mechanism for the formation of inter-mediates leading to the formation of SAICAR.

2. Materials and methods

2.1. Expression, purification and characterization of *sp*PurC

All reagents used in buffers and for cell growth were purchased from Thermo Fisher Scientific (Waltham, Massachusetts, USA) unless otherwise noted. *Escherichia coli* BL21-Gold (Stratagene, Santa Clara, California, USA) cells with the *sp*PurC gene (Accession No. L15190; Hui & Morrison, 1993) in pET-15b vector, containing an N-terminal His tag, were grown overnight to give a starter culture (30 ml). This culture was then used to inoculate 2 l Luria Broth medium containing 0.1 mg ml^{-1} ampicillin and was grown at 37°C for 3 h with shaking before the addition of isopropyl β-D-1-thiogalacto-pyranoside (Gold Biotechnology, St Louis, Missouri, USA) to a final concentration of 0.5 m*M*. The culture was allowed to grow for an additional 4 h and the cells were harvested and stored at −20°C. The frozen cells (5 g) were suspended at 4°C in 60 ml of 50 m*M* tris(hydroxymethyl)aminomethane (Tris) buffer with 500 m*M* NaCl and 10 m*M* imidazole at pH 8.0 (binding buffer) plus 1% Triton X-100, 250 µl protease-inhibitor cocktail (Sigma–Aldrich, St Louis, Missouri, USA) and 5 mg lysozyme (Sigma–Aldrich). The cells were then subjected to sonication on ice. The soluble fraction was collected by centrifugation and loaded onto a column of nickel metal-affinity resin (Qiagen, Valencia, California, USA) with binding buffer. His-tagged *sp*PurC was released from the column using an elution buffer (binding buffer plus 500 m*M* imidazole), concentrated and loaded onto a Superdex 200 column *via* an ÄKTA FPLC (GE Life Sciences, Pittsburgh, Pennsylvania, USA) for further purification and for hydrodynamic mass analysis. The purified protein sample (with a 20-residue His tag at the N-terminus followed by the 235-residue *sp*PurC) was frozen dropwise in liquid nitrogen and stored at −80°C. The protein identity was first verified by gel electrophoresis and then by mass spectrometry (high-resolution, high mass accuracy measurements with an LTQ-FT spectrometer at the Research Resources Center, University of Illinois at Chicago).

2.2. Crystallization

All crystallization reagents, plates and related tools were purchased from Hampton Research (Aliso Viejo, California, USA) unless otherwise noted. CAIR was synthesized

according to previously published methods (Mehl & Begley, 2002). The thawed His-tagged *sp*PurC (350 µ*M*) was dialyzed overnight at 4°C in 15 m*M* Tris buffer pH 8.0 containing 25 m*M* KCl, 55 m*M* MgCl$_2$, 5 m*M* ethylenediaminetetraacetic acid and 5 m*M* dithiothreitol, followed by centrifugation to remove any precipitate. High-throughput crystal-growth screening was carried out at the Research Resources Center, University of Illinois at Chicago utilizing a Tecan Freedom Evo liquid-handling robot and the screening solutions, PEGs Suite and JCSG+ Suite of Qiagen NeXtal DW Block at 16°C for a week before crystals were examined. The final optimized well solution [0.16 *M* sodium acetate pH 4.5, 36–40%(*w/v*) PEG 200] was used to produce *sp*PurC crystals under two different conditions: condition 1 contained 5 m*M* ADP (MP Biomedicals, Solon, Ohio, USA), 1 m*M* CAIR and 4 m*M* Asp (Thermo Fisher Scientific), while condition 2 contained only 10 m*M* ADP. Vapor diffusion by the hanging-drop method was utilized with 2 µl 250 µ*M* protein solution and 2 µl well solution.

2.3. Data collection, processing, structure solution, refinement and analysis

The crystals were directly flash-cooled in a liquid-nitrogen stream without cryoprotectant. A single crystal from each condition was used for diffraction, and data were collected using a MAR 225 detector on the Southeast Regional Collaborative Access team (SER-CAT) 22-ID beamline at the Advanced Photon Source at Argonne National Laboratory (Argonne, Illinois, USA). The X-ray source was set to a wavelength of 1.000 Å. 400 frames for structure 1, using a crystal from condition 1, and 360 frames for structure 2, using a crystal from condition 2, were collected with a scan width of 0.5°, an exposure time of 1 s and a crystal-to-detector distance
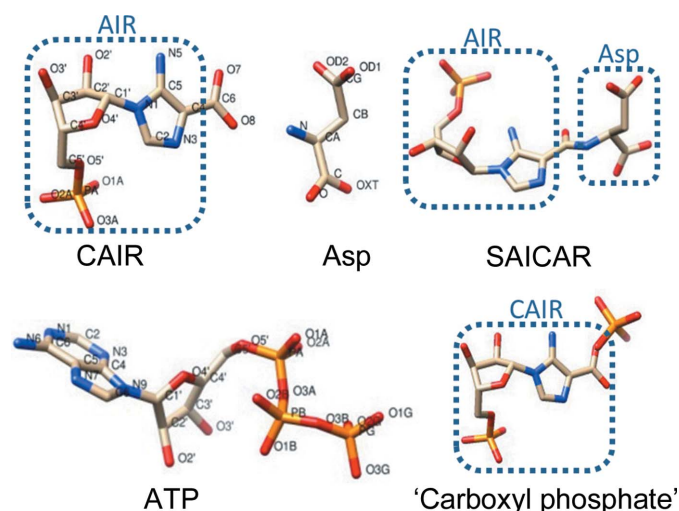
Figure 1
The structure and the numbering of the atoms of each molecule involved in the PurC-catalyzed reaction. CAIR, Asp, SAICAR and ATP are from X-ray structures in the PDB: PDB entries 2gqs, 4fe2, 2cnv and 1obd, respectively. The structure of 'carboxyl phosphate' was assembled from the structures of CAIR and the γ-phosphate of ATP.

Table 1
Data-collection and refinement statistics for *sp*PurC structure 1 (PDB entry 4fe2) and structure 2 (PDB entry 4nye).

Values in parentheses are for the highest resolution shell.

Structure	1 (PDB entry 4fe2)		2 (PDB entry 4nye)	
Chain	A	B	A	B
Space group	$P2_1$		$P2_1$	
Unit-cell parameters (Å, °)	$a = 46.07$, $b = 65.02$, $c = 86.05$, $\beta = 92.09$		$a = 45.93$, $b = 65.11$, $c = 85.84$, $\beta = 91.96$	
Resolution (Å)	40.00–2.30 (2.34–2.30)		40.44–2.69 (2.74–2.69)	
Multiplicity	3.7 (3.6)		2.7 (2.6)	
Mosaicity (°)	1.3		1.6	
No. of unique reflections	21601		10808	
No. of atoms per asymmetric unit	4356		4025	
Completeness (%)	98.5 (86.8)		82.7 (69.5)	
R_{merge} (%)	8.1 (19.0)		9.2 (18.5)	
$\langle I/\sigma(I)\rangle$	25.9 (9.3)		11.8 (8.3)	
R factor	0.196 (0.204)		0.229 (0.321)	
R_{free}	0.284 (0.297)		0.314 (0.380)	
Wilson B factor (Å^2)	23.3		33.9	
R.m.s.d. from ideal values				
Bond lengths (Å)	0.016		0.011	
Bond angles (°)	2.023		1.667	
Ramachandran plot				
Preferred residues	434 [93.5%]		418 [90.1%]	
Allowed residues	25 [5.4%]		34 [7.3%]	
Outlier residues	5 [1.1%]		12 [2.6%]	
Ligands in active site	ADP, Asp, AIR, Mg^{2+}	ADP, AIR, Mg^{2+}	ADP, acetate, Mg^{2+}	ADP, acetate, Mg^{2+}
No. of waters	210	204	86	75
Visible residues	3–235	1–235	1–235	2–235
Mean B factors (Å^2)				
Protein	25.2	28.3	41.3	44.1
Water	33.5	36.3	39.9	39.7
ADP	22.7	25.3	38.8	40.5
Asp	49.6			
AIR	28.2	28.4		
Mg^{2+}	30.3	27.8	48.1	52.2
Cl$^-$	23.0		42.2	
Acetate		43.4	40.3	54.4

of 250 mm. The *HKL*-2000 software package was used for processing (Otwinowski & Minor, 1997).

Structure 2 was solved using molecular replacement with the *MOLREP* software as implemented in the *CCP*4 package (Winn *et al.*, 2011), utilizing the structure of *E. coli* PurC (*ec*PurC) with ADP and Mg^{2+} (PDB entry 2gqr; Ginder *et al.*, 2006) as a search model. The initial solution had a correlation coefficient of 0.414 (solved using data in the resolution range 40–4.4 Å) and an *R* factor of 0.54. Amino-acid sequence revision and subsequent rigid-body refinement confirmed the solution, with an R_{free} of 0.41 and R_{work} of 0.31. Although a reasonable data set was collected to 2.6 Å resolution for this structure, the final refinement was performed only using data to 2.7 Å resolution. The extent and the completeness (84.4%; 69.5% in the last resolution shell; Table 1) of the data for this structure were sub-optimal for a thorough refinement; however, since the available data for the last resolution shell were strong [$\langle I/\sigma(I)\rangle$ = 8.3; Table 1], the crystallographic refinement was pursued with tighter geometrical restraints (see the values for the r.m.s.d. for bond lengths and angles, the *R* factor and the R_{free} in Table 1). Structure 1 was then solved using structure 2 for molecular replacement to obtain a starting model for refinement with a correlation coefficient of

0.639 and an *R* factor of 0.409. For both structures, additional model revisions and addition of solvent and ligands for the final rounds of refinement were completed using *Coot* (Emsley *et al.*, 2010). Data-collection and final refinement statistics are presented in Table 1.

Structure analysis and presentation were performed with *PyMOL* (Schrödinger, New York, USA) and *UCSF Chimera* v.1.8.0 (Pettersen *et al.*, 2004). An OMIT $2F_o - F_c$ electron-density map in the vicinity of the active site in structure 1 chain *A* contoured at 1.0σ was calculated after removing ADP, Asp and AIR from the model and performing three cycles of refinement with *Coot* (Emsley *et al.*, 2010) to validate the ligand electron density. Dimer-interface analysis was performed with *PDBsum* (Laskowski, 2009). Ligand–protein interatomic distances were obtained with either *LPC/CSU* (Sobolev *et al.*, 1999; http://ligin. weizmann.ac.il/cgi-bin/lpccsu/LpcCsu.cgi) or *PyMOL*. The active-site volume was estimated with *Pocket-Finder* (http:// www.modelling.leeds.ac.uk/pocketfinder/). 81 PurC gene sequences in the National Center for Biotechnology Information database CDD v3.10 (http://www.ncbi.nlm. nih.gov/Structure/cdd/wrpsb.cgi) were compared to identify conserved residues in these genes. The reaction mechanism in Fig. 5(*b*) was drawn with *ChemBioDraw* (v.13.0.2, Perkin Elmer).

2.4. Modeling the active site of *sp*PurC with all substrates

We modeled the active-site structure of *sp*PurC with the complete set of ligands (ATP, Mg^{2+}, Asp, CAIR and waters) needed for the formation of SAICAR. ATP was found in PurC from the thermophilic archaeon *Methanocaldococcus jannaschii* (*mj*PurC; PDB entry 2z02; RIKEN Structural Genomics/Proteomics Initiative, unpublished work) and in PurC from the yeast *Saccharomyces cerevisiae* (*sc*PurC; PDB entry 1obd; Levdikov *et al.*, 1998). CAIR was found in the Gram-negative *E. coli* PurC (*ec*PurC; PDB entry 2gqs; Ginder *et al.*, 2006). The all-atom root-mean-square deviation (r.m.s.d.) value of our (*sp*PurC) structure 1 chain *A* to the monomeric *sc*PurC is 2.28 Å and that to chain *A* of *mj*PurC is 0.64 Å. The larger r.m.s.d. between *sp*PurC and *sc*PurC is owing to insertions in the sequence of *sc*PurC. The r.m.s.d. of *sp*PurC and *ec*PurC is 0.72 Å. Thus, these four structures are relatively similar to each other. The ATP in *mj*PurC has the γ-phosphate 'bent down' to hydrogen-bond to its own ribose group (bent ATP conformation) and the ATP in *sc*PurC

extends 'out' (linear ATP conformation). The ATP molecules of scPurC and mjPurC superimposed on the ADP of spPurC with r.m.s.d.s of 0.20 and 0.36 Å, respectively. CAIR of ecPurC superimposed on AIR of spPurC with an r.m.s.d. of 0.10 Å.

To create our models, we started with chain A of structure 1 of spPurC (see below) harboring ADP, Asp and AIR. We replaced the ADP with the ATP from either mjPurC chain A (bent ATP conformation) or scPurC (linear ATP conformation) and the AIR with the CAIR from ecPurC. We first removed the Mg^{2+}, if present, from all structures. We also either added or replaced some water molecules. All waters hydrogen-bonded to ligands in the original structures were retained. For ATP with the bent conformation, the γ-phosphate was rotated to orient towards Mg^{2+} and the side chain
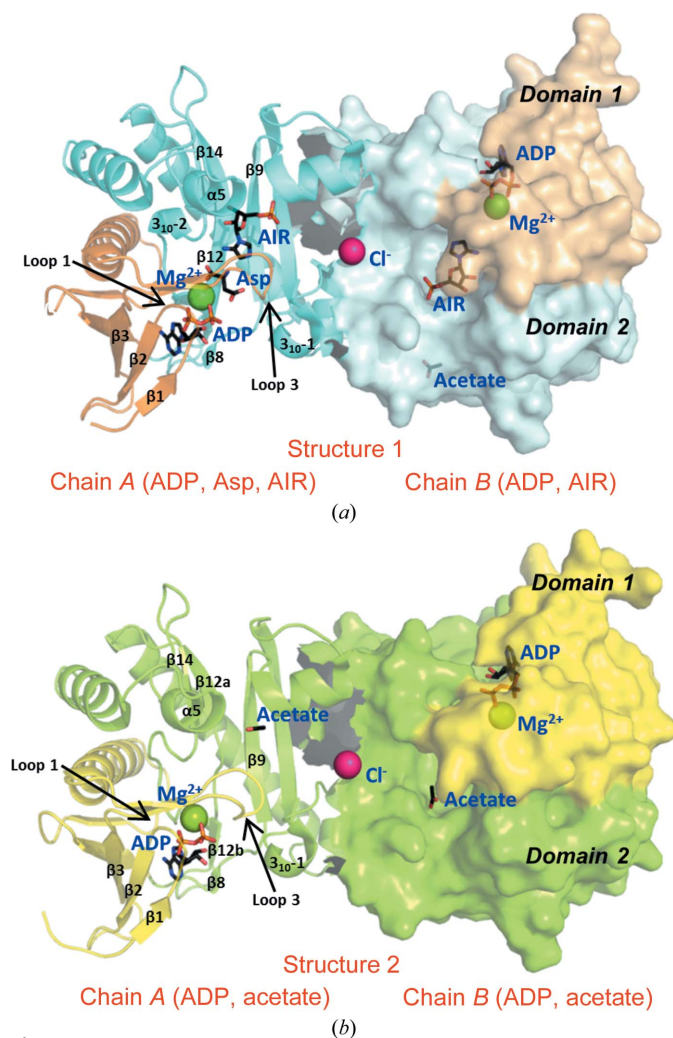
Figure 2
Structures of the spPurC dimer with (a) ADP/AIR/Asp (PDB entry 4fe2) or (b) ADP (PDB entry 4nye) in the active sites. Chain A is depicted in a cartoon representation with domain 1 (residues 1–81) in orange and domain 2 (residues 82–235) in aqua. Chain B is depicted with surface detail. Ligand molecules are shown in black with selected atoms colored (N in blue, O in pink and P in orange). Mg^{2+} ions are shown in green and the Cl^- ion (at the dimer interface) is shown in magenta. The names of the ligands are indicated in blue and the secondary-structural elements are indicated in black. Residues 171–181 form a long β12 strand in the presence of AIR (structure 1) or β12a, an unstructured region and β12b in the absence of AIR (structure 2).

Table 2
Summary of the secondary-structural elements of spPurC structure 1 chains A and B.

The secondary-structural features of these two chains are very similar. A second 3_{10}-helix was present in chain A but not in chain B.

	Residues†
β-Strand	
1	5–8
2	12–16
3	22–27
4	30–32/31
5	39/40–41
6	69–72
7	77–81
8	84–85
9	88–95
10	111–121
11	130–131
12	171–181
13	187–190
14	198–201
3_{10}-Helix	
1	123–127
2	194–196‡
α-Helix	
1	44–63
2	97–105/104
3	132–139
4	143–167/168
5	210–215
6	219–232
Loop	
1	9–11
2	17–21
3	33–38/39
4	64–68
5	73–76
6	106/105–110
7	182–186
8	191–193/197
9	202–209

† The number before the slash is the residue number of chain A and the number after the slash is the residue number of chain B when chain B differs from chain A. ‡ This 3_{10}-helix is present in chain A only.

of Lys122 was replaced with a rotamer from the Dunbrack library within *UCSF Chimera*. To neutralize charges, Mg^{2+} ions were added to coordinate the ATP and the carboxylate group of CAIR. After the initial model setup, energy minimization of each structure was achieved with the Molecular Modelling Toolkit provided in *UCSF Chimera*, using default settings to achieve an energy minimum.

3. Results

About 200 mg of His-tagged spPurC protein was purified from 5 g of cells to >90% purity. The molecular mass from the mass-spectrometric analysis was 29 034.9 Da, suggesting that the initial methionine residue in the His-tag portion was cleaved *in situ*, since the expected value for the sequence without the His-tag methionine residue is 29 034.8 Da. The hydrodynamic mass of spPurC was found to be about 69 kDa, indicating that it was dimeric in solution. Crystals of spPurC appeared as thin diamond-shaped plates of about 30–50 μm in the largest dimension. The monoclinic crystal belonged to space group

$P2_1$ and contained a homodimer (chains A and B) in the asymmetric unit. Structure 1 was prepared using spPurC in the presence of ADP, CAIR, Asp, Mg^{2+} and acetate, which resulted in both chains A and B containing ADP, AIR and Mg^{2+} in the active site. Interestingly, Asp was also found only in structure 1 chain A. Structure 2 was prepared using spPurC in the presence of ADP, Mg^{2+} and acetate, all of which were found in both active sites (Fig. 2b). The resolution was 2.3 Å for structure 1 and 2.7 Å for structure 2. Although structure 2 was a lower resolution structure (see discussion in §2.3), it allowed comparison with structure 1 to determine any possible major conformational changes upon the binding of the ligands AIR and Asp.

3.1. General structure of spPurC

For the dimer (chain A and chain B) of structure 1 (PDB entry 4fe2) 468 residues were resolved, which did not include the residues in the His-tag segment in both chains and the first two residues in chain A (Table 1). The Ramachandran plot showed 94% of the residues in the preferred region, with 5% of the residues in the allowed region and 1% of the residues (mostly in the flexible loop 3) in the outlier region. We found an AIR molecule, instead of CAIR (Fig. 1), in the crystal prepared with CAIR (Figs. 2, 3 and 4). CAIR, although stable, has been reported to decompose to AIR in acidic solution, with ring protonation promoting the removal of the carboxylate group at the C_4 position (Fig. 1; Litchfield & Shaw, 1971), and our crystals were formed in sodium acetate solution.

Each chain of spPurC folded into domains 1 and 2, following the previously published nomenclature (Ginder *et al.*, 2006), with residues 1–81 (domain 1) consisting of $\beta1$–$\beta7$

(a)

(b)

(c)
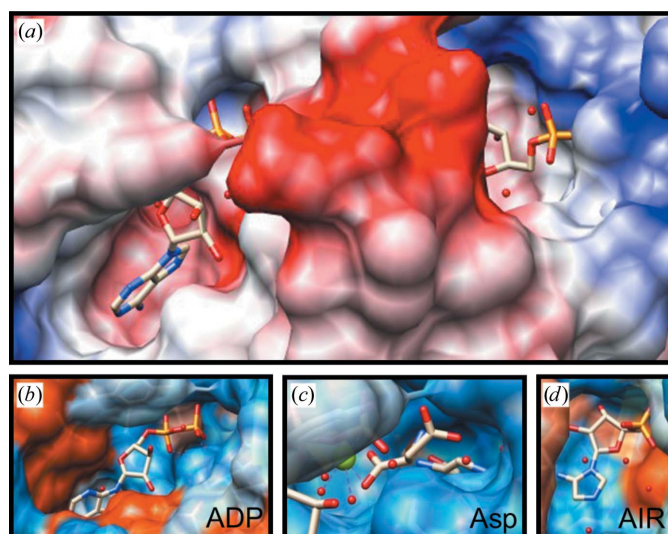
Figure 3
Electrostatic representation of the active-site tunnel of structure 1 chain A, with positive charges in red and negative charges in blue. Generally, the active-site tunnel is highly negatively charged, with positive charges at each exit (a). The solvent-exposed portion of loop 3 is seen as positively charged folding over at the center. The adenine ring of ADP is at the left side of the tunnel (b) and the phosphate of AIR is at the right side of the tunnel (d). The Asp pocket, enclosed by loop 3 at the top, is highly negatively charged (c).

(a)

(b)

(c)
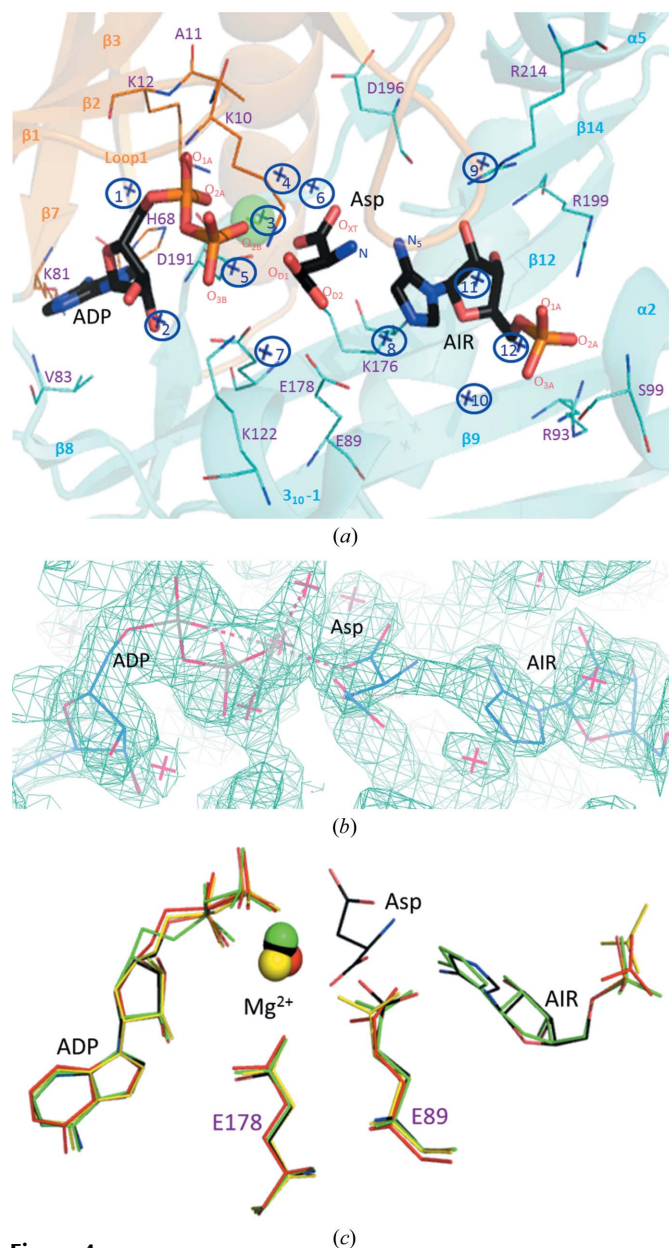
Figure 4
(a) The active site of structure 1 (PDB entry 4fe2) chain A with ADP, Mg^{2+}, Asp, AIR and 12 water molecules (marked as blue crosses) shows that the side chains of residues Lys10, Lys12, Lys81 and Asp191 interact with ADP. Lys10, Ala11 and Lys12 are also involved in backbone-atom interactions. The side chains of residues Lys10, Lys122 and Lys176 interact with Asp. The side chains of residues Glu89, Arg93, Ser99, Asp196, Arg199 and Arg214 interact with AIR. A network of water molecules, 1–12 (representing waters 406, 517, 504, 505, 506, 535, 571, 555, 402, 412, 452 and 495 in the X-ray structure, respectively), is also found to be involved in interactions with side chains and ligands. The backbone is depicted with transparency to highlight the side chains. (b) An OMIT $2F_o - F_c$ electron-density map in the vicinity of the active site in structure 1 chain A, contoured at 1.0σ, confirmed density for ADP, Asp and AIR within the active site. The positions of the ligands in all four active sites are very similar (c). Ligands in structure 1 chain A (ADP, Mg^{2+}, Asp and AIR) are in black as in (a), those in structure 1 chain B (ADP, Mg^{2+} and AIR) are in green, those in structure 2 chain A (ADP) are in red and those in structure 2 chain B (ADP) are in yellow. Most of the side-chain positions remain similar, including the proposed catalytic glutamate residues. AIR can be seen to shift slightly out of the tunnel to accommodate Asp. The acetate molecules of structure 2 are seen near the phosphate group of AIR.

and α1 (see Table 2 for the residues in β-strands, helices and loops in these structures), and residues 82–235 (domain 2) consisting of β8–β14, α2–α6 and a 3_{10}-helix (Fig. 2, Table 2). The two chains in structure 1 were generally similar (Table 2), with an all-atom r.m.s.d. value of 0.17 Å. Residues 194–196 formed a 3_{10}-helix in chain A but were in an unstructured region in chain B. The B-factor values were highest for loop 3 (residues Ala33–Lys39).

For structure 2 (PDB entry 4nye) 469 residues were resolved (Table 1). The two chains were again similar, with an all-atom r.m.s.d. value of 0.32 Å. The secondary-structural elements were similar to those in structure 1 (Table 2), except for β12. The β12 strand found in structure 1 was no longer a long β-strand consisting of residues 171–181, but was split into two β-strands, β12a (residues 171–175 in chain A and 170–177 in chain B) and β12b (residues 179–181 in chain A and 179–182 in chain B), connected by unstructured residues (Fig. 2). Structures 1 and 2 were similar, with an all-atom r.m.s.d. value of 0.30 Å.

The detailed structural features of the active site will be discussed below. Outside the active site, an acetate ion was found in structure 1 chain B near a surface residue in β9: Asn94. In both structures 1 and 2, a $Cl^−$ ion was found at the interface of the dimer, interacting with residue Arg103 in helix α2 of both chains. When all atoms of the two structures were superimposed, the two $Cl^−$ ions were at very similar positions, about 0.2 Å apart.

The interfaces of the dimers in both structures were comprised of 17 residues from each monomer, with surface areas of 1790 $Å^2$ for structure 1 and 1740 $Å^2$ for structure 2. In structure 1, ten hydrogen bonds (involving residues Arg103, Val117, Glu118, Phe130, Asn132, Glu134, His135 and Arg155) and 111 nonbonding contacts were found. In structure 2, five hydrogen bonds (involving residues Arg103, Val117, Glu118, Phe130, Glu134 and His135) and 110 nonbonding contacts were found.

We also found several Tris molecules decorating different parts, but not the active site, in structure 1. Three of the Tris molecules were found in chain A near α1, α4 and β10, while a fourth molecule was found near the 3_{10}-helix.

3.2. Active-site structures

The active site of spPurC in all structures was relatively large (1495 and 1752 $Å^3$ for structure 1 chains A and B, respectively), at approximately 7% of the total volume (23 047 and 23 193 $Å^3$ for structure 1 chains A and B, respectively), and was comprised of a hydrophilic tunnel between domains 1 and 2, with a network of hydrogen bonds between water molecules, side chains and ligand molecules binding to the active sites (Fig. 3). Close examination of the active site of structure 1 chain A, which contained the most complete set of ligands, revealed details of this deep tunnel, providing the binding sites for ADP, Mg^{2+}, Asp and AIR (Figs. 3 and 4). The OMIT electron-density map (Fig. 4b) confirmed the presence of these ligands. β12 was at the bottom of the tunnel. One wall of the tunnel (wall 1) was formed by a β-sheet consisting of β1–β3 and loop 1 in domain 1 and β14, 3_{10}-helix 2 and α5 in domain 2 (Fig. 2a). The other wall (wall 2) was formed by β8 and β9 in domain 2. The top of the two walls merged to form the ceiling of the tunnel with parts of domain 1 (loop 3) and domain 2 (3_{10}-helix 1), which is highly negatively charged. ADP, Mg^{2+}, Asp and AIR generally interacted with charged residues and a network of water molecules in this active-site tunnel (Fig. 4a).

ADP was at one end, above β12, with the adenine ring near the opening of the tunnel, interacting with hydrophobic residues such as Ala11 and Ile85 and with His68 (Figs. 3 and 4). The N_6 atom (following the PDB nomenclature for atoms; Fig. 1) in the adenine ring of ADP was hydrogen-bonded to His68, and N_7 was hydrogen-bonded to Asp191 and a water molecule, water molecule 1 (or 406 in structure 1 chain A; Fig. 4). The phosphate groups interacted with the N atoms of the side chains and the backbone of several lysine residues (Lys10 in loop 1, Lys12 in β2 and Lys122 in β10), as well as with water molecules 2–5 (or 517 and 504–506) and Mg^{2+} (Fig. 4). The α- and β-phosphate groups as well as the three water molecules 3–5 (or 504–506) interacted with the hexa-oxygen-coordinated Mg^{2+} ion, and the sixth coordination was to the O atom of the terminal carboxylate group of Asp (O_{XT}).

The conformation of the Asp molecule was a $gauche(−)$ conformation, with a distorted side-chain χ_1 torsion angle of $−24°$ about the $N—C_A—C_B—C_G$ bond, presumably owing to repulsion from the nearby Glu89 residue side chain. Asp interacted with Lys176 (β12) and Lys10 (loop 1) and AIR as well as four water molecules: 3 and 6–8 (or 535, 571 and 555). Interacting atoms and distances are shown in Table 3.

AIR, in addition to the interaction of its N_5 atom with Asp, also interacted with Arg93 (with the phosphate of AIR), Ser99 (with the phosphate), Asp174 (with the ribose and phosphate), Asp196 (with the ribose), Arg199 (with the phosphate) and Arg214 (with the ribose) as well as water molecules 9–12 (or 402, 412, 452 and 495) (Table 3). Water molecule 8 also interacted with AIR at the N_3 position. Between Asp and AIR was an opening that was large enough to accommodate a carboxylate group at C_4 of AIR to give the CAIR molecule.

The other three active sites in our structures were generally similar to that of structure 1 chain A (Fig. 4c), but without the Asp in structure 1 chain B and without Asp and AIR for those in structure 2 (PDB entry 4nye). Superimposing all atoms of the four chains showed that the positions of all four ADPs were very similar, with each corresponding atom pair differing by <0.8 Å in the adenine rings, <0.6 Å in the ribose rings and <0.2 Å for the P atoms (Fig. 4c). The phosphate group rotated about the phosphate ester bond, resulting in the corresponding O atoms being about 1 Å apart. With structure 1 chain A as a reference, the phosphate ester bond pairs of structure 1 chain B were rotated clockwise, and in structure 2 both chains were rotated counterclockwise by about 10–15°. Thus, the binding of Asp to the active site appeared to cause only minor changes in the positions and structures of ADP, AIR and Mg^{2+} in the active site. Asp and/or AIR binding caused a slight movement of the nearby residue Asn35 in loop 3. The active site in structure 1 chain B was covered by loop 3.

Table 3
Binding contacts in structure 1 of the *sp*PurC structure with ADP, Asp and AIR, as well as contacts of ATP, Asp and CAIR in the bent ATP model.

Contacts are defined as less than or equal to 3.6 Å.

Ligand	Partner‡	Distance† (Å)		
		Chain A	Chain B	Model
ADP/ATP				
N_1	Val83 N	3.2	3.1	3.1
N_3	W1675 O			3.4
N_6	Lys81 O	2.8	2.8	2.8
	His68 N_{D1}	2.9	3.0	3.1
N_7	Asp191 N	3.3	3.2	3.3
	W505 O	3.4		
	W406 O	3.4	3.4	
$O_{2'}$	Glu178 O_{E2}	2.9		2.9
	Glu178 O_{E1}		2.9	
$O_{3'}$	W517 O	3.0		2.8
$O_{5'}$	Lys12 N_Z			3.6
O_{1A}	Ala11 N	2.6	2.8	2.8
	Lys12 N	2.7	2.8	2.8
	Lys10 N	3.1	3.0	3.1
O_{2A}	Mg^{2+}–ATP	2.3		1.9
	Lys12 N_Z	2.8	3.0	3.1
	Asp191 O_{D1}	3.6		
	W504 O	3.1		2.6
	W506 O	2.9		
	W505 O			2.9
O_{1B}	Mg^{2+}–ATP	1.7		2.2
	Asp O_{D1}	3.0		3.0
	W505 O	3.1	3.4	2.9
	W506 O	3.0		
	W509 O		3.1	
O_{2B}	Mg^{2+}–ATP	2.1		
	Lys10 N	2.9	3.1	2.7
	Lys10 N_Z	3.6		
	W475 O		2.6	
O_{3B}	W517 O	2.7		3.2
	Lys122 N_Z	2.8		
	W520 O		3.1	
O_{1G}	Glu178 O_{E1}			3.2
	Mg^{2+}–ATP			1.9
	Asp O_{XT}			2.9
	W504 O			3.0
	W1624 O			2.8
O_{3G}	Lys122 N_Z			2.7
	W571 O			2.8
	W1624 O			2.6
Mg^{2+}_{CAIR}	Asp128 O_{D2}			2.8
	Glu89 O_{E1}			1.9
	W248 O			2.3
	W249 O			2.0
Mg^{2+}_{ATP}	W505 O	2.2	2.4	1.9
	W506 O	2.2	2.2	
	W509 O		2.4	
	W504 O	2.3	2.0	1.9

Table 3 (continued)

Ligand	Partner‡	Distance† (Å)		
		Chain A	Chain B	Model
AIR/CAIR				
N_3	Glu89 O_{E1}		3.5	3.4
	W555 O	2.6	2.6	
	W452 O	3.6		
	Mg^{2+}–CAIR			2.9
	W249 O			2.9
	W248 O			3.5
N_5	Asp196 O	3.4	3.1	3.3
	W262 O			3.3
O_7	W262 O			2.7
	W452 O			3.4
O_8	Mg^{2+}–ATP			1.7
	W248 O			2.7
	W249 O			3.3
$O_{2'}$	Arg214 N_{H2}	2.9	2.8	2.8
	Asp196 O	3.5	3.2	3.6
	Arg214 N_{H1}		3.5	3.5
	W402 O	2.8	3.0	
$O_{3'}$	Asp174 O_{D1}	2.3	2.6	2.4
	Arg199 N	3.5		3.4
$O_{5'}$	Arg199 N_{H1}	3.4	3.2	
O_{1A}	W495 O	2.8		2.8
	W452 O	3.3		3.4
	Arg199 N_{H2}	3.6	3.4	
	W487 O		2.3	
O_{2A}	Arg199 N_{H1}	2.9	3.1	2.7
	Arg93 N_{H1}	3.0	2.7	3.5
	Ser99 N	3.0	3.1	3.0
	Arg199 N_{H2}			3.2
	Arg93 N_{H2}	3.3	3.3	
O_{3A}	W412 O	2.8	2.6	3.4
	Ser99 O_G	3.0	3.2	2.5
	Arg93 N_{H2}	3.2	3.1	3.1
	W495 O	3.3		
Asp				
N	AIR N_5	3.1		
	CAIR O_7			2.6
O	CAIR N_5	3.1		3.1
	W535 O	2.5		3.1
	Lys176 N_Z			3.6
O_{XT}	Lys176 N_Z	2.5		3.1
	Mg^{2+}–ATP	2.6		2.1
	W535 O	2.8		3.2
	W505 O			2.7
	W504 O			2.6
	W1624 O			2.5
O_{D1}	Lys10 N_Z	2.7		2.8
	W571 O			3.4
O_{D2}	Lys10 N_Z			3.6
	W555 O	3.0		
	W571 O	3.1		

† Any redundant measurements are presented only once. ‡ Water is abbreviated as 'W' throughout the table.

Lys39 in this loop interacted with Asp122 in the neighboring crystal contact, bringing loop 3 closer to the active site. In addition, the side chain of the Asn35 residue was nearer to the active site, allowing the environment of the various residues in the vicinity of the active site of chain B to differ from those in chain A. Presumably owing to these differences, we observed the ligand Asp in chain A but not in chain B in structure 1. This loop was very flexible: it moved by the greatest distance between chains and retained the highest B factors. In structure 2, without AIR in the active site, an acetate ion was found in the vicinity of the phosphate group of AIR, and in the absence of Asp the Lys176 side chain was in the location of the imidazole ring of AIR. The movement of Lys176 split the long β-strand (β12) in structure 1 into β12a and β12b connected by a disordered region in structure 2 (Fig. 2*b*).

3.3. A proposed active-site structure with ATP, Mg^{2+}, Asp and CAIR

Structure 1 chain A of *sp*PurC (the 'parent' active site), although being the closest representation of the enzyme–

substrate complex of any PurC to date, did not contain two important functional groups: the γ-phosphate of ATP and the carboxylate group of CAIR. In our proposed structure of the active site with a full set of ligands, the γ-phosphate of ATP, with either a 'bent' or a 'linear' conformation (see §2), was positioned near the base or the top of the active-site tunnel, respectively. For the bent conformation, the side chain of Lys122, which is in an unstructured region adjacent to β10, was rotated to avoid clashes with the γ-phosphate of ATP (Fig. 5a). In the linear conformation, the γ-phosphate of ATP points towards the top of the active-site tunnel, near Lys10, Ala11 and Lys12 in loop 1 (structure not shown). The ADP part of the ATP interacted with side chains similar to those found in the 'parent' active site. The distances between interacting atoms are listed in Table 3.

With ATP in either the bent or linear conformation, the amino group of Asp rotated towards the top of the active-site tunnel, while the carboxyl group of CAIR rotated with respect to the imidazole ring of CAIR to avoid clashes with the amino group of Asp and to allow a hydrogen-bond interaction of CAIR O_7 and Asp H_3, and was no longer in the planar conformation exhibited in ecPurC. A similar non-planar conformation has been observed in X-ray structures containing 5-aminoimidazole-4-carboxamide 1-β-D-ribofuranoside (AICAR; Day et al., 2007; Jin et al., 2007; Zhang, White et al., 2008).

4. Discussion

Several structures of PurC from bacteria (both Gram-positive and Gram-negative), archaea and eukaryota, as well as the PurC domain of the bifunctional protein phosphoribosyl-aminoimidazole carboxylase (PAICS) from human and silkworm have been solved, and some have been published (Ginder et al., 2006; Levdikov et al., 1998; Li et al., 2007; Manjunath et al., 2013; Taschner et al., 2013; Urusova et al., 2003, 2006; Zhang et al., 2006). These structures are generally similar. The structures of our chains A and B were similar to each other, and were similar to the monomer structure of PurC of *Clostridium perfringens*, which belongs to the same phylum (Firmicutes), and also to that of *Mycobacterium abscessus* (phylum Actinobacteria) as well as those of the Gram-negative *E. coli* and *Ehrlichia chaffeensis*. The active sites of

these structures usually contain various relevant molecules and ions (ADP, Mg^{2+} and molecules mostly from the crystal-growth solution; Ginder et al., 2006), but none contained the substrate Asp. Thus, none of the active sites contained a full set of ligands (ATP, Mg^{2+}, Asp and CAIR). Owing to their lability, CAIR and AIR have been termed 'elusive compounds' (Groziak et al., 1988). Based on our structure (structure 1 chain A) with the active site consisting of Asp, as well as ADP, Mg^{2+} and AIR, we modeled ATP and CAIR into the active site.

The ADPs in the four active sites in our two structures (PDB entries 4fe2 and 4nye) and the ADPs or ATPs in other PurC structures (PDB entries 3nua, 2gqr, 2gqs, 2ywv, 2yzl, 2z02, 2cnq, 1obd and 1obg) are all in about the same position.
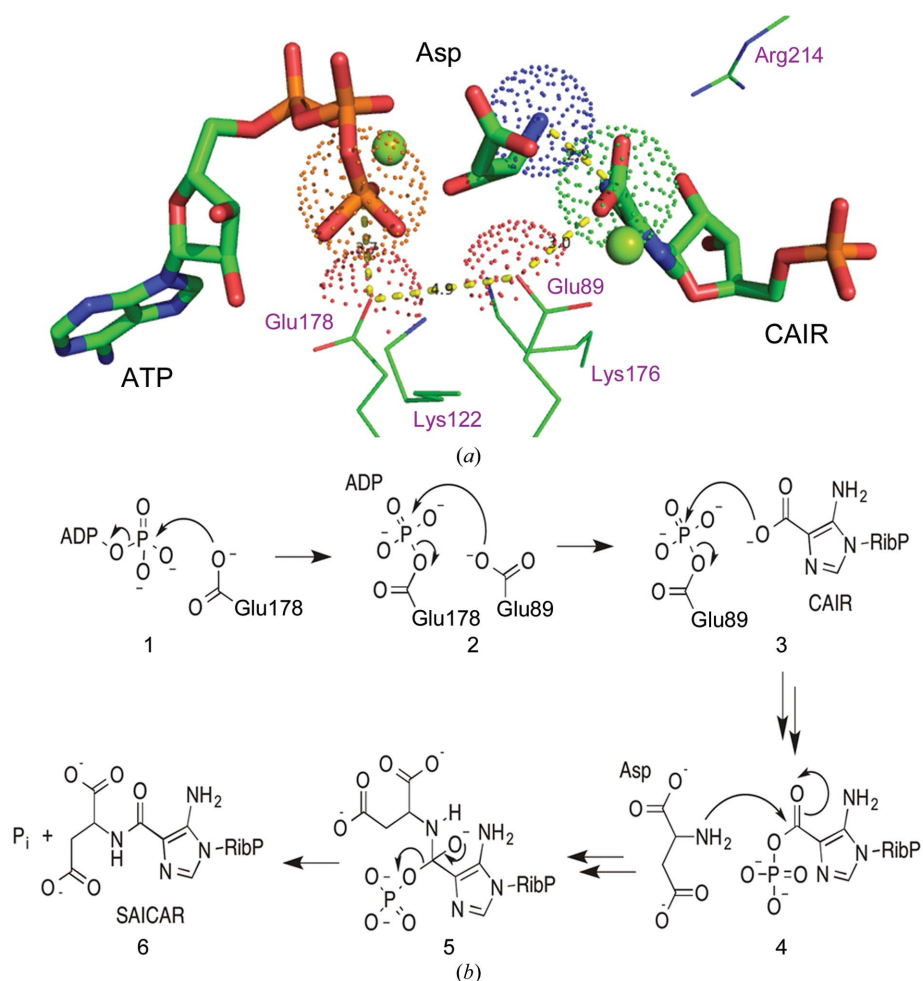
Figure 5
(a) The model structure of the active site with ATP, Mg^{2+}, Asp, CAIR and a network of water molecules after energy minimization suggests that Glu178, Glu89, Lys122 and Lys176 are important residues in the reaction leading to SAICAR. The ATP is shown to have a 'bent' conformation. The van der Waals radii for the important groups (the γ-phosphate of ATP, the N atom of Asp, C_6 of the carboxylate of CAIR, Glu178 O_{E1} and Glu89 O_{E2}) are also shown. (b) Relay process: the γ-phosphate of ATP and the Asp178 O_{E1} atoms involved in the first step of the mechanism are within van der Waals distances. The proposed reaction mechanism for PurC is initiated by a side-chain carboxylate attack of the γ-phosphate of ATP, freeing ADP. The phosphate is subsequently transferred to a second carboxylate and then to CAIR. After removal of the hydrogen of the amino group of Asp, the N-Asp attacks CAIR to form a tetrahedral intermediate, which will decompose to SAICAR after intramolecular hydrogen removal from the amino group of Asp and loss of phosphate. Double arrows represent a deprotonation event at the amino group of Asp. RibP represents a 5'-phosphoribosyl group.

The CAIR in *ec*PurC (PDB entry 2gqs) and the AIR in *sp*PurC are also in about the same position. Therefore, we believe that the structures of our model active site with ATP, Mg^{2+}, Asp and CAIR are the most plausible models and thus are suitable for detailed examination for mechanistic insights.

The mechanism of condensation between the C$_4$ carboxylate group (in the –COO$^-$ form) of CAIR and Asp should be very similar to that of the ATP-dependent coupling of carboxylate and amino groups in which the carboxylate is activated by phosphorylation *via* transfer of the terminal phosphate group of ATP, leading to the formation of an activated ester (Shomura *et al.*, 2012). The mixed phosphoric anhydride of CAIR (Fig. 1) then acrylates the amino group of Asp to generate SAICAR and phosphate. The mechanism of this proposed process is quite similar to that of ATP-dependent synthetases and ligases (Frey & Hegeman, 2007). Since the ATP and CAIR in our model structures are at the distal ends of the active-site tunnel, direct activation of the carboxylate of CAIR by ATP, as suggested previously (Ginder *et al.*, 2006; Nelson *et al.*, 2005), appears to be improbable. We also believe that the reaction mechanism proposed previously (Ginder *et al.*, 2006), involving initial reaction between Asp and CAIR to form a dioxyanion intermediate followed by ATP phosphorylation, is improbable owing to the low reactivity of the carboxylate towards nucleophiles.

It is most likely that the activation of the carboxylate of CAIR is a relay process initiated by phosphate transfer to the carboxylate of a conserved side chain (1 in Fig. 5*b*; Smith *et al.*, 2009). For ATP in the bent conformation, the carboxylate group of the Glu178 side chain is within van der Waals distance of the γ-phosphate (Fig. 5*a*), allowing the γ-phosphate to be transferred to a second carboxylate in Glu89 and finally to the carboxylate of CAIR (2 and 3 in Fig. 5*b*). The presumably protonated amino group of Asp could be deprotonated by the phosphate group or by a water molecule bound between the amino group of CAIR and the carboxylate of Asp, as found in the structure of *ec*PurC. After proton transfer, the free amino group of Asp then attacks the activated carboxylate group in CAIR, forming a tetrahedral intermediate (4 and 5 in Fig. 5*b*), which then expels the phosphate group through elimination with concomitant proton transfer, generating SAICAR and a phosphate molecule (5 and 6 in Fig. 5*b*). It should be mentioned that the side-chain conformation of Asp may exhibit several other positions to that presented here to facilitate the carboxylate activation of CAIR to form the phosphoric anhydride intermediate. Our structure, however, is the only structure with Asp in the active site; thus, we are not able to predict other variations at this time. The proposed mechanism for the bent ATP conformation model involves several residues (Glu89, Lys122, Lys176 and Glu178) which are conserved in 79 of the 81 known PurC gene sequences. The negative charges on the released phosphate would be stabilized by the Lys122 and Lys176 side chains, as well as by the Mg^{2+} ions. A water molecule in the active site, found bound to CAIR in *ec*PurC (water 262) and mentioned above, may play an important role. This water is hydrogen-bonded to the conserved Arg214,

which may also stabilize the phosphate group in the tetrahedral intermediate.

For the structure with ATP in the linear conformation, without substantial backbone movement, we found no side chains with carboxylate groups within the van der Waals radii of the γ-phosphate, with the closest being that of Asp196, which was 6 Å away. A less energetically favorable mechanism could involve a water coordinated to Mg^{2+} becoming deprotonated by Asp196, allowing the resulting hydroxyl group to hydrolyze the γ-phosphate. The free phosphate would then need to travel towards CAIR to activate the carboxylate of CAIR. This appeared to be a less likely process.

We based our modeling of ATP in the active site with the two available X-ray structures, the bent and linear conformations, and favored the bent conformation. However, the position of the γ-phosphate in ATP–Mg^{2+} is known to be quite variable and flexible (Liao *et al.*, 2004; Antonyuk *et al.*, 2001). Thus, it is possible for the γ-phosphate to rotate and then assume a favorable position to facilitate the phosphorylation of the Glu89 side chain to start the relay process of carboxylate activation.

It is interesting to note that part of the active site of the PurC domain in human PAICS (hPur6; PDB entry 2h31; Li *et al.*, 2007) is also quite similar to that of *sp*PurC except in the ATP-binding site; more specifically, in the loop 1 position. However, the top portion of the active site (β4, loop 4, β5, loop 9 and α5 in *sp*PUrC) is missing in the X-ray structure (Li *et al.*, 2007), presumably owing to flexibility. If this is the case, these differences in the active sites may provide the specificity needed to design inhibitors of bacterial PurC but not of the PurC domain of human PAICS.

5. Conclusion

The structure of *sp*PurC with ADP, AIR and Asp represents the only published structure of PurC containing most relevant ligands in the active site and is the first one with Asp in the active site. This structure provides details about the binding of each ligand. Small secondary-structural differences can be observed with and without AIR and Asp bound. These changes include movement in a flexible loop (loop 3), a conformational change in β12 containing residue Lys176 owing to AIR binding, and the formation of a second 3$_{10}$-helix at Asp196 induced by Asp binding. However, in general, the structures with and without the ligands, and the structures across many bacterial species, are quite similar. We believe that our model active-site structure including ATP–Mg^{2+}, Asp, AIR and a network of water molecules is quite realistic and is well suited to a detailed examination of the reaction mechanism. We propose a relay process for carboxylate activation to transfer the γ-phosphate from ATP to the side chains of conserved residues (Glu89 and Glu178). The negative charges of the ultimately released phosphate would be stabilized by the side chains of, again, conserved residues (Lys122 and Lys176). Future studies with mutations of these residues will further our understanding of PurC reactions. Thus, this

study provides the basic framework for the structure-based rational design and development of PurC inhibitors.

References

Antonyuk, S. V., Grebenko, A. I., Levdikov, V. M., Urusova, D. V., Melik-Adamyan, V. R., Lamzin, V. S. & Wilson, K. S. (2001). *Crystallogr. Rep.* **46**, 620–625.

Cornick, J. E. & Bentley, S. D. (2012). *Microb. Infect.* **14**, 573–583.

Day, P., Sharff, A., Parra, L., Cleasby, A., Williams, M., Hörer, S., Nar, H., Redemann, N., Tickle, I. & Yon, J. (2007). *Acta Cryst.* D**63**, 587–596.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Frey, P. A. & Hegeman, A. D. (2007). *Enymatic Reaction Mechanisms*, pp. 547–548. Oxford University Press.

Ginder, N. D., Binkowski, D. J., Fromm, H. J. & Honzatko, R. B. (2006). *J. Biol. Chem.* **281**, 20680–20688.

Groziak, M. P., Bhat, B. & Leonard, N. J. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 7174–7176.

Hui, F. M. & Morrison, D. A. (1993). *J. Bacteriol.* **175**, 6364–6367.

Jin, X., Townley, R. & Shapiro, L. (2007). *Structure*, **15**, 1285–1295.

Laskowski, R. A. (2009). *Nucleic Acids Res.* **37**, D355–D359.

Levdikov, V. M., Barynin, V. V., Grebenko, A. I., Melik-Adamyan, W. R., Lamzin, V. S. & Wilson, K. S. (1998). *Structure*, **6**, 363–376.

Li, S.-X., Tong, Y.-P., Xie, X.-C., Wang, Q.-H., Zhou, H.-N., Han, Y., Zhang, Z.-Y., Gao, W., Li, S.-G., Zhang, X. C. & Bi, R.-C. (2007). *J. Mol. Biol.* **366**, 1603–1614.

Liao, J.-C., Sun, S., Chandler, D. & Oster, G. (2004). *Eur. Biophys. J.* **33**, 29–37.

Litchfield, G. J. & Shaw, G. (1971). *J. Chem. Soc. B*, pp. 1474–1484.

Lynch, J. P. III & Zhanel, G. G. (2010). *Curr. Opin. Pulm. Med.* **16**, 217–225.

Manjunath, K., Kanaujia, S. P., Kanagaraj, S., Jeyakanthan, J. & Sekar, K. (2013). *Int. J. Biol. Macromol.* **53**, 7–19.

Mehl, R. A. & Begley, T. P. (2002). *J. Labelled Comp. Radiopharm.* **45**, 1097–1102.

Nelson, S. W., Binkowski, D. J., Honzatko, R. B. & Fromm, H. J. (2005). *Biochemistry*, **44**, 766–774.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.

Poll, T. van der & Opal, S. M. (2009). *Lancet*, **374**, 1543–1556.

Samant, S., Lee, H., Ghassemi, M., Chen, J., Cook, J. L., Mankin, A. S. & Neyfakh, A. A. (2008). *PLoS Pathog.* **4**, e37.

Shomura, Y., Hinokuchi, E., Ikeda, H., Senoo, A., Takahashi, Y., Saito, J., Komori, H., Shibata, N., Yonetani, Y. & Higuchi, Y. (2012). *Protein Sci.* **21**, 707–716.

Smith, A. J. T., Li, Y. & Houk, K. N. (2009). *Org. Biomol. Chem.* **7**, 2716–2724.

Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E. & Edelman, M. (1999). *Bioinformatics*, **15**, 327–332.

Taschner, M., Basquin, J., Benda, C. & Lorentzen, E. (2013). *Proteins*, **81**, 1473–1478.

Urusova, D. V., Antonyuk, S. V., Grebenko, A. I., Lamzin, V. S. & Melik-Adamyan, V. R. (2003). *Crystallogr. Rep.* **48**, 763–767.

Urusova, D. V., Levdikov, V. M., Antonyuk, S. V., Grebenko, A. I., Lamzin, V. S. & Melik-Adamyan, V. R. (2006). *Crystallogr. Rep.* **51**, 824–827.

Winn, M. D. et al. (2011). *Acta Cryst.* D**67**, 235–242.

Zhang, R., Skarina, T., Evdokimova, E., Edwards, A., Savchenko, A., Laskowski, R., Cuff, M. E. & Joachimiak, A. (2006). *Acta Cryst.* F**62**, 335–339.

Zhang, Y., Morar, M. & Ealick, S. E. (2008). *Cell. Mol. Life Sci.* **65**, 3699–3724.

Zhang, Y., White, R. H. & Ealick, S. E. (2008). *Biochemistry*, **47**, 205–217.