Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Elucidation of the bicarbonate binding site and insights into the carboxylation mechanism of (N^5)-carboxyaminoimidazole ribonucleotide synthase (PurK) from *Bacillus anthracis*

Micheal L. Tuntland,[a] Bernard D. Santarsiero,[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

Received 21 January 2014
Accepted 23 September 2014

PDB references: PurK, 4dlk; 3v4s; 3qff; 3r5h

Structures of (N^5)-carboxyaminoimidazole ribonucleotide synthase (PurK) from *Bacillus anthracis* with various combinations of ATP, ADP, Mg^{2+}, bicarbonate and aminoimidazole ribonucleotide (AIR) in the active site are presented. The binding site of bicarbonate has only been speculated upon previously, but is shown here for the first time. The binding involves interactions with the conserved residues Arg272, His274 and Lys348. These structures provide insights into each ligand in the active site and allow a possible mechanism to be proposed for the reaction that converts bicarbonate and AIR, in the presence of ATP, to produce (N^5)-carboxyaminoimidazole ribonucleotide. The formation of a carboxyphosphate intermediate through ATP phosphoryl transfer is proposed, followed by carboxylation of AIR to give the product, facilitated by a cluster of conserved residues and an active-site water network.

1. Introduction

The *de novo* purine-biosynthetic pathway converts 5-phospho-α-D-ribosyl-1-pyrophosphate to inosine monophosphate, from which adenine and guanine are derived. This conversion is a six-enzyme, ten-step process in higher eukaryotes, including humans, but is a ten-enzyme, 11-step process in lower eukaryotes and prokaryotes such as *Escherichia coli* and *Bacillus anthracis* (Kappock *et al.*, 2000; Zhang *et al.*, 2008). The sixth step in the higher eukaryote pathway is carried out by phosphoribosylaminoimidazole carboxylase synthetase (PAICS, also known as Pur6) as a single bifunctional enzyme to convert aminoimidazole ribonucleotide (AIR) and CO_2 to carboxy-amino-imidazole ribonucleotide (CAIR) (Patey & Shaw, 1973). In the prokaryotic pathway there are two steps, with the first step being the reaction of bicarbonate (HCO_3^-) and AIR in the presence of adenosine triphosphate (ATP) to produce (N^5)-carboxyaminoimidazole ribonucleotide (N^5-CAIR) by N^5-CAIR synthase (PurK) (Firestine *et al.*, 1994). Subsequently, N^5-CAIR mutase (PurE) converts N^5-CAIR, which is quite unstable, to CAIR (Meyer *et al.*, 1999). In fungi, PurK and PurE are fused together as a bifunctional enzyme to convert AIR and bicarbonate to N^5-CAIR and then N^5-CAIR to CAIR (Firestine *et al.*, 1998; Thoden *et al.*, 2010). PurK enzymes have been targeted for the development of antifungal drugs (Firestine *et al.*, 1998) and antibacterial drugs (Brugarolas *et al.*, 2011). In humans, mutations in the enzymes of this pathway lead to the accumulation of intermediates, resulting in hereditary disorders such as AICA-ribosiduria (Marie *et al.*, 2004) and autism (Sivendran *et al.*, 2004). However, the carboxylation mechanisms involving HCO_3^- in the prokaryotic pathway and CO_2 in the eukaryotic pathway are not well

Table 1
X-ray data and refinement statistics.

Values in parentheses are for the highest resolution shell.

PDB code	3r5h†	3v4s‡	3qff§	4dlk¶
Crystal parameters				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	*a* = 57.44, *b* = 82.79, *c* = 168.23	*a* = 57.74, *b* = 82.87, *c* = 167.93	*a* = 57.42, *b* = 82.09, *c* = 166.77	*a* = 57.37, *b* = 84.48, *c* = 167.27
Data collection				
Resolution range (Å)	19.76–2.20 (2.26–2.20)	19.96–2.02 (2.14–2.02)	19.97–1.96 (2.08–1.96)	19.99–2.02 (2.14–2.02)
No. of unique reflections	41312	51528	55699	52033
Average multiplicity	6.0 (5.5)	5.5 (3.8)	7.6 (5.1)	4.5 (2.9)
Completeness (%)	99.4 (94.9)	96.1 (77.2)	97.2 (82.1)	95.6 (78.8)
$\langle I/\sigma(I)\rangle$	22.6 (8.7)	22.4 (5.5)	14.4 (2.0)	14.5 (2.9)
$R_{\rm merge}$††	0.054 (0.199)	0.047 (0.206)	0.091 (0.816)	0.075 (0.415)
Wilson *B* factor (Å^2)	35	28	39	21
Refinement				
No. of protein residues	760	761	754	760
No. of water molecules	557	483	317	308
Chain *A*				
No. of AIR molecules	1	0	0	0
No. of HCO_3^- molecules	0	1	0	0
No. of Mg atoms	1	1	0	0
No. of Ca atoms‡‡	0	0	0	3
No. of ATP molecules	0	0	0	3
No. of ADP molecules	1	1	1	0
Chain *B*				
No. of AIR molecules	1	0	0	0
No. of HCO_3^- molecules	0	1	0	0
No. of Mg atoms	1	2	0	0
No. of Ca atoms‡‡	0	0	0	2
No. of ATP molecules	0	1	0	1
No. of ADP molecules	1	0	1	0
$R_{\rm work}$§§	0.179 (0.201)	0.175 (0.208)	0.199 (0.312)	0.198 (0.262)
$R_{\rm test}$¶¶	0.248 (0.270)	0.217 (0.260)	0.252 (0.389)	0.235 (0.292)
Average *B* factor (Å^2)	29.2	31.3	34.7	24.3
Clashscore†††	2.4	3.7	2.7	1.5
Poor rotamers†††	11	6	3	3
R.m.s. deviations from ideal‡‡‡				
Bond lengths (Å)	0.011	0.008	0.010	0.006
Bond angles (°)	1.31	1.26	1.18	1.15
Ramachandran plot†††§§§				
Favorable	731 [96.7%]	742 [98.1%]	729 [97.2%]	744 [98.4%]
Acceptable	25 [3.3%]	14 [1.9%]	20 [2.7%]	11 [1.5%]
Outliers	0 [0.0%]	0 [0.0%]	1 [0.1%]	1 [0.1%]

† The crystal-growth conditions for 3r5h were as follows. The well solution consisted of 0.1 *M* MES, 5 m*M* DTT, 10 m*M* $MgCl_2$, 11% glycerol, 10% PEG 6000 (pH 6.5). Hampton Research Additive Screen condition No. 46, which consists of 100 m*M* ATP (0.5 µl), was added to the protein drop (2.0 µl protein solution and 0.5 µl well solution) to give a final concentration of 16.7 m*M* ATP. Crystals were soaked in 10 m*M* AIR for 30–60 s prior to flash-cooling for diffraction measurements. ‡ For 3v4s, the same well solution as above was used except that it also contained 100 m*M* $NaHCO_3$. ATP (100 m*M*, 0.5 µl) was added to the protein drop (2.0 µl protein solution and 0.5 µl well solution). § For 3qff, the well solution consisted of 0.1 *M* MES, 5 m*M* DTT, 9% glycerol, 10% PEG 6000 (pH 6.5). Hampton Research Silver Bullets Bio C3/27 (0.5 µl) was added to the protein drop (1.0 µl protein solution and 0.5 µl well solution). The crystal was soaked in 3 m*M* ADP for 30–60 s prior to flash-cooling. ¶ For 4dlk, a solution consisting of 0.1 *M* MES, 5 m*M* DTT, 10 m*M* $Ca(NO_3)_2$, 9% glycerol, 9% PEG 6000 (pH 6.5) and ATP (100 m*M*, 0.25 µl) was added to the protein drop (1.0 µl protein solution and 0.25 µl well solution). †† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡‡ Ca^{2+} was used to assist in the determination of the position(s) of the Mg^{2+} ion since the electron density of Mg^{2+} and water may appear similar. §§ $R_{\rm work} = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}||/\sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure factors, respectively. ¶¶ $R_{\rm test}$ is calculated in the same way as $R_{\rm work}$ but using test reflections that were randomly selected and excluded from the refinement. ††† The values obtained were given by *MolProbity* (Chen *et al.*, 2010). ‡‡‡ The values are in comparison to those of Engh & Huber (1991). §§§ The values are given as percentages in brackets.

understood. Previously reported structures of PurK include those from *E. coli* (*ec*PurK), with Protein Data Bank (PDB; Berman *et al.*, 2000) codes 1b6r, 1b6s, 3eth and 3etj (Thoden, Kappock *et al.*, 1999; Thoden *et al.*, 2008), from *Aspergillus clavatus* (*ac*PurK; PDB entries 3k5i and 3k5h; Thoden *et al.*, 2010) and from *Staphylococcus aureus* (*sa*PurK; PDB entries 3orq and 3orr; Brugarolas *et al.*, 2011), as well as our previous structure of PurK from *B. anthracis* (*ba*PurK) with only Mg^{2+} ions in the active site (Tuntland *et al.*, 2011). We now report structures of *ba*PurK with each of the ligands needed for the reaction in the active site, *i.e.* ATP-Mg^{2+}, bicarbonate, AIR and a water network, although not all are present together in the active site at the same time. These structures provide the first details of the bicarbonate binding site in PurK. Owing to the similarity found in the active-site structures when different ligands are bound, we are able to model an active-site structure including all of the ligands at the same time. This model allows us to propose mechanistic insights into the carboxylation step of an essential pathway, *de novo* purine biosynthesis, in *B. anthracis*.

2. Materials and methods

2.1. Cloning, expression and purification

The *ba*PurK gene was cloned, expressed and purified as described by Tuntland *et al.* (2011). Protein samples were frozen dropwise in liquid nitrogen and stored at 193 K until needed.

2.2. Crystallization

The protein solution was concentrated to 26 mg ml^{-1} (725 µ*M*) in 10 m*M* Tris buffer pH 8.0 with 200 m*M* NaCl. Crystallization was accomplished using the hanging-drop vapour-diffusion method (Tuntland *et al.*, 2011). All reservoir (well) solutions consisted of 0.1 *M* MES buffer pH 6.5, 9–11% glycerol, 9–10% PEG 6000 and DTT. Some wells also contained $NaHCO_3$. The final concentration of ATP, when present, was 16.7 m*M* in all of the drops. Various ligand solutions were added to the protein drops or soaked into the crystals prior to flash-cooling and diffraction measurements. Detailed conditions for specific crystals are given in the footnotes to Table 1.

2.3. Data collection and processing, and structure determination

Data were collected on SER-CAT beamline 22-ID (Southeastern Regional Collaborative Access Team, Advanced Photon Source, Argonne National Laboratory) using a 300 mm MAR CCD detector for image collection. *XDS* (Kabsch, 2010) was used for indexing, integration and scaling. Coordinates from the previously published *ba*PurK structure with only Mg^{2+} in the active site (PDB entry 3q2o; Tuntland *et al.*, 2011) were used as the starting model for the structures obtained in this study. *Coot* (Emsley & Cowtan, 2004) was used for ligand insertion and manual rebuilding, and *REFMAC* (Murshudov *et al.*, 2011) was used for refinement in *CCP*4 (Winn *et al.*, 2011). Explicit details of each refined structure are given in Table 1. *MolProbity* (Chen *et al.*, 2010) was used for structure validation prior to deposition in the PDB. Structural images were generated using *PyMOL* (Schrödinger). The r.m.s.d. from *VMD* and *PyMOL* was used to compare distance fluctuations (Humphrey *et al.*, 1996). The active-site volume was determined with *Pocket-Finder* (http://www.modelling.leeds.ac.uk/pocketfinder) and the value was rounded for appropriate precision consideration (Novotny *et al.*, 2007). Sequences of homologous enzymes/domains were obtained using the *BLASTP* tool from the Network Protein Sequence Analysis server. Sequence alignment and comparison was carried out using *Jalview* v.2.6.1 (Waterhouse *et al.*, 2009) and *ClustalW* (Larkin *et al.*, 2007).

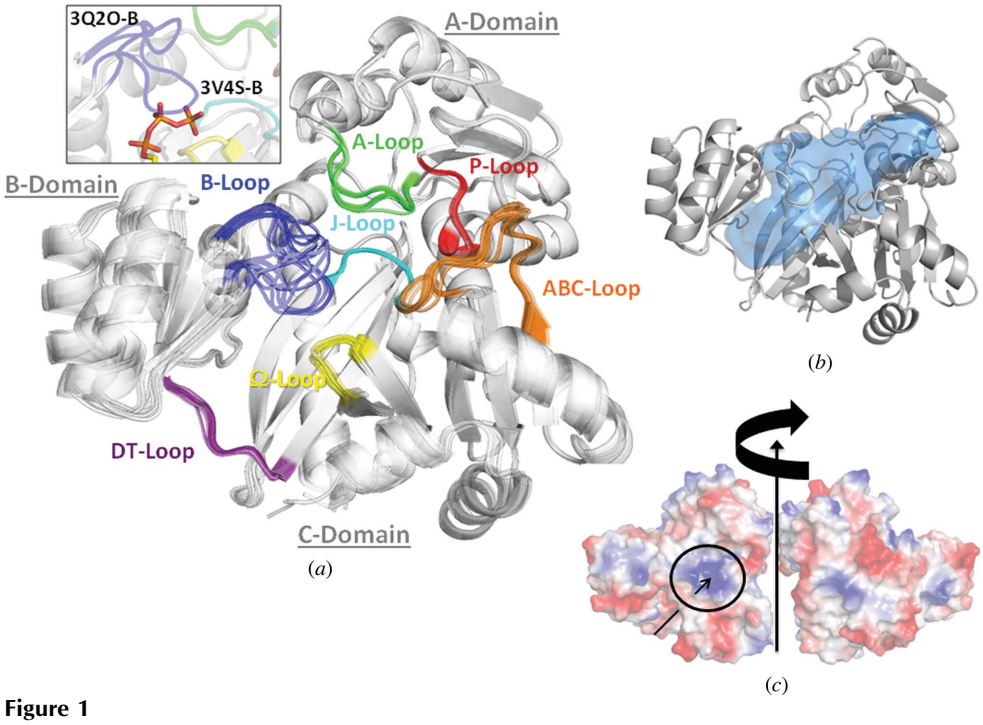

Figure 1
(*a*) Overlay of all eight chains from the four structures, as well as the two chains of the previously published structure (PDB entry 3q2o) of *ba*PurK. Most of the structural elements, including the major ligand-interacting loops, the J-loop (cyan), P-loop (red) and Ω-loop (yellow), exhibit little movement, with the exceptions being the A-loop (green), the B-loop (blue) and the ABC-loop (orange). The B-loop is the most flexible and is in an 'open' conformation or 'up' position without ATP but is in a 'closed' conformation or 'down' position with ATP (see top inset; the 'open' conformation is from 3q2o chain *B* and the 'closed' conformation is from 3v4s chain *B*). (*b*) The volume of the active site with ATP, Mg^{2+} and HCO_3^- (3v4s chain *B*) is approximately 1900 Å^3, as generated by the *Pocket-Finder* server, and is shown as a blue space-filled pocket. (*c*) The electrostatic potential of the sites for AIR, HCO_3^- and the phosphates of ATP is highly positive (blue), while the rest of the surface is mostly negative (red) or neutral (white). The arrow shows the active-site tunnel. The tail of the arrow shows the site to which the adenine ring of ATP binds and the head of the arrow shows the site to which AIR binds, with HCO_3^- in between ATP and AIR. The circled region consists of a cluster of positive charges in the active site. A 180° rotation about the vertical axis shows that the rest of the enzyme surface is relatively less charged than the active site.

3. Results

*ba*PurK is a dimer in solution (Tuntland *et al.*, 2011). The samples of *ba*PurK with different ligands all crystallized in space group $P2_12_12_1$ with two independent chains (*A* and *B*), often with different ligand(s) within each chain in the asymmetric unit. The structures solved from four data sets are shown in Fig. 1(*a*) and Table 1. The resolutions of the structures range from 1.96 to 2.20 Å. We found all of the reaction substrates, AIR, HCO_3^-, ATP-Mg^{2+} and a water network, not necessarily all together at the same time, in the active sites of two structures (3r5h chain *B* with AIR and 3v4s chain *B* with HCO_3^- and ATP-Mg^{2+}). Other

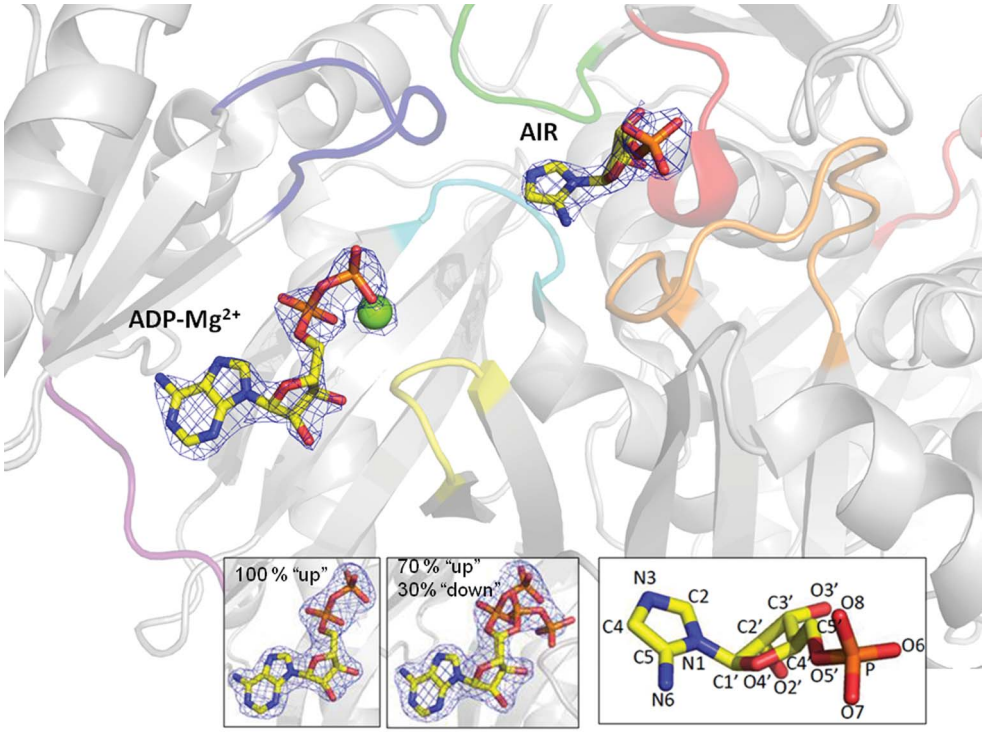

Figure 2
ADP-Mg^{2+} and AIR in the active site of *ba*PurK (PDB entry 3v4s chain *B*) are shown. AIR has not been previously observed in the active site of a bacterial PurK. The electron-density map of ADP [a $2mF_o - DF_c$ map at 2.0σ (0.61 e Å^{-3}) using *CCP*4 and *PyMOL*] shows ADP in the 'normal' conformation, the same as that in active sites with only ADP and no Mg^{2+} (3qff chain *A*, left inset; 3qff chain *B*, middle inset). In the middle inset, only 70% of the ADP is in this normal 'up' position. 30% of ADP is observed with the β-phosphate pointing 'down'. The electron-density maps in the insets are shown at 1.0σ (0.27 e Å^{-3}). The lower right inset shows the structure and atom names of AIR, with the exocyclic amine N atom shown as the sixth atom (N6) in the imidazole ring according to the X-ray software nomenclature. This atom is termed N^5 in the text and in the literature (Mueller *et al.*, 1994).

ligands, such as ADP, were also found in the active sites of some of these structures (Table 1). In general, the active-site structures, other than the positions of some loops (discussed below), are similar and are relatively large and open. For example, the volume of the active site of 3v4s chain *B* is about 1900 Å^3 or about 5% of the total enzyme volume (Fig. 1*b*), with many polar and positively charged residues (Fig. 1*c*) to accommodate the highly negatively charged substrates ATP, HCO_3^- and AIR. The details of the active-site structures are discussed below. The relatively large active site is able to accommodate three ATP molecules with Ca^{2+} in one crystal (4dlk chain *A*). Calcium was used to assist in the assignment of the position(s) of Mg^{2+} ion in the active site. The electron density of Mg^{2+} and water may appear similar, but generally the coordination geometry is octahedral for cations and tetrahedral for water. The coordination geometry for both Mg^{2+} and Ca^{2+} was octahedral. Nitrate, the counter-ion of Ca^{2+}, was not observed in 4dlk, although nitrate and bicarbonate have similar geometries. Outside the active site, a lone PO_4^{3-} ion was found around residues 46–50 of both the *A* and *B* chains within the interface of one structure (4dlk).

*ba*PurK, as a member of the ATP-grasp superfamily, consists of three domains (A, B and C; Tuntland *et al.*, 2011) with four common loops: the P-loop (residues Gly15–Leu19), the B-loop (Thr149–Gly157), the Ω-loop (Val214–Ile217) and the J-loop (Ala270–His274) (Tuntland *et al.*, 2011; Thoden, Kappock *et al.*, 1999). In this study, we identified three additional loops that are involved in interactions with the active-site ligands (Fig. 1). The first loop, a six-residue loop (Tyr75–Ile80) interacting with AIR *via* Glu76, is designated the A-loop (for **A**IR). The second loop, another six-residue loop (Lys183–Glu188) interacting with ADP/ATP, is designated the DT-loop (for A**D**P/A**T**P). Lastly, a 12-residue loop (Leu337–Lys348) interacting with both AIR and bicarbonate is designated the ABC loop.

All of the structures are similar to each other, except for the positions of some loops (Fig. 1*a*), and are similar to the structure with only Mg^{2+} cations in the active site (PDB entry 3q2o; Tuntland *et al.*, 2011). The C^α r.m.s.d. for each of the eight chains to 3q2o chain *A* is <0.85 Å. Most of the secondary-structural elements, including the P-loop, Ω-loop and J-loop, exhibit little movement upon binding ligands. However, the A-loop, B-loop, ABC-loop and DT-loop exhibit differing degrees of movement upon binding ligand(s), with the glycine-rich B-loop (T^{149}TGGYDGKG^{157}) being the most flexible. In other ATP-grasp enzymes, this loop is considered to be a 'flexible lid' over the active site (Fawaz *et al.*, 2011; Thoden, Wesenberg *et al.*, 1999; Ye *et al.*, 2001). In the absence of the terminal phosphate of ATP, the B-loop exhibits an 'open conformation' or 'up position' (for example, as in 3q2o

chain *B* with no ATP; Fig. 1*a*, inset), but is in a 'closed conformation' or 'down position' when ATP is present (3v4s chain *B* and 4dlk chain *B*).

3.1. ADP/ATP binding

For the first time, ATP-Mg^{2+} was found with HCO_3^- in the active site (3v4s chain *B*), with a network of water molecules in close proximity around the γ-phosphate and HCO_3^- (Supplementary Table S1[1]). The binding of HCO_3^- is discussed below.

In the active sites of the four structures we found either ADP or ATP, and they are in approximately the same location at one end of the active site with or without Mg^{2+}, HCO_3^- and/or AIR. The ADP part of ATP (3v4s chain *B*) is in approximately the same position as ADP (3r5h chain *B*) in the active site. As noted above, the flexible B-loop exhibits an 'up position' in the absence of the terminal phosphate of ATP (Fig. 1*a*, inset), but in the presence of ATP shows a 'down position' (3v4s chain *B* and 4dlk chain *B*). The positions of this loop in the active sites of structures with other ligands are in between these two extreme positions (Fig. 1*a*). Upon binding ATP, some residues in the loop, including Gly152, Tyr153 and Gly155, move 5–10 Å (see Supplementary Table S1 for the distances between interacting atoms). The side-chain interactions include Gln158 (the residue immediately following the B-loop) with the α-phosphate, Glu190 with the ribose moiety and Arg107, Lys147 and Asn267 with phosphate groups. The hydrated Mg^{2+} ion interacts with the terminal phosphate to stabilize the orientation and position of ADP/ATP. For ADP-Mg^{2+} (3r5h chain *A*, 3r5h chain *B* and 3v4s chain *A*), the β-phosphate is 'pointing up' (Fig. 2). However, in the absence of Mg^{2+} the β-phosphate can either be 'pointing up' with 100% occupancy (3qff chain *A*) or can have a dual occupancy with 70% 'pointing up' and 30% 'pointing down' (3qff chain *B*). This 'down' conformation interacts with the same residues as the 'up' conformation, as well as with the side chains of His213 and Glu255. The backbone atoms of the DT-loop (K^{183}WVPFE^{188}) residues Lys183 and Val185 are in contact with the adenine ring. In addition, the side chains of Trp184, Phe187 and Phe257 are involved in a π–π interaction with the adenine ring (Supplementary Table S1). A number of water molecules, such as water 375 from 3v4s chain *B*, are also found in each of the active sites and display a coordination geometry typical of water and not that of metal cations. One of the two Mg^{2+} ions found in structure 3v4s chain *B* is in the same position as one Ca^{2+} ion in the structure 4dlk chain *B*, both coordinating to the phosphate groups of ATP. Furthermore, two additional ATP molecules, one occupying the HCO_3^- site and another the AIR site, were found in one structure (4dlk chain *A*). Recently, it has been shown that ATP (up to 5 m*M*) inhibits the enzymatic reaction in *ac*PurK and it has been proposed that multiple ATP molecules can occupy the active site (Dewal & Firestine, 2013). Indeed, the active site may accommodate three ATP molecules, at least for *ba*PurK.

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: KW5089).

3.2. AIR binding

The electrostatic potential of the binding site for AIR, adjacent to the ATP-binding site (Fig. 2), is relatively positive (3r5h chains *A* and *B*; Fig. 1*b*). The ribose of AIR (see the inset in Fig. 2 for the structure and naming of the AIR atoms) interacts with Glu76 in the A-loop, and the phosphate of AIR interacts with Lys340 and Arg347 in the long ABC-loop (Supplementary Table S1). Lys340 and Arg347 as well as Lys348 (which interacts with HCO_3^-; see below) are conserved across 12 different species, including archaea, bacterial prokaryotes (both Gram-positive and Gram-negative) and the lower eukaryotes fungi and plants (Supplementary Fig S1). In addition, AIR also interacts with Gln18 and Leu19 in the P-loop. The location of AIR in the active site of *ac*PurK (PDB entry 3k5i, with ADP-Mg^{2+} and AIR in the active site; Thoden *et al.*, 2010) is nearly the same as in *ba*PurK. The residues Gln14, Glu73, Lys345 and Arg352 that interact with AIR in *ac*PurK correspond to Gln18, Glu76, Lys340 and Arg347 in *ba*PurK. Residue Phe77 in the A-loop of *ba*PurK appears to behave differently to Phe78 of *sa*PurK (Brugarolas *et al.*, 2011), in that the Phe78 side chain occupies the location of AIR when AIR is absent (Supplemetary Table S1). However, in *ba*PurK Phe77 is positioned away from the active site at about the same location with or without AIR in the active site. As part of the important water network, water 479 of 3r5h chain *B* is hydrogen-bonded to N^5 of AIR (at a distance of 2.6 Å).

3.3. Bicarbonate binding

HCO_3^- has been proposed to bind in the space between ATP and AIR (Dewal & Firestine, 2013), but the location of HCO_3^- in the active site has not been reported previously. This study reveals for the first time the specific location of HCO_3^- in the active site of *ba*PurK. HCO_3^-, with a unique trigonal planar density, is equidistant between ATP and AIR (3v4s chains *A* and *B*; Fig. 3). Since HCO_3^- has not been observed previously, we rigorously examined our density assignment. Three different OMIT maps were generated: $mF_o - DF_c$, $2mF_o - DF_c$ and $3mF_o - 2DF_c$ (Supplementary Fig. S2). After removing HCO_3^- from the structural model and re-refining the structure, each of these maps still generated a trigonal planar-shaped patch of density at 3.5σ (0.28 e Å^{-3}), 1.5σ (0.43 e Å^{-3}) and 2.5σ (0.78 e Å^{-3}), respectively. An additional $mF_o - DF_c$ OMIT map (Supplementary Fig. S3) was generated as above with residues Arg272 and Lys348 and the HCO_3^- ion omitted from the model to confirm our placement of HCO_3^- and to measure its interactions (Supplementary Table S1). In *E. coli* PurK structures other ions are found in this location (*e.g.* a sulfate ion in 1b6r and a phosphate ion in 3etj; Thoden, Kappock *et al.*, 1999; Thoden *et al.*, 2008). Aligning the C^α atoms in 1b6r and in 3etj with our 3v4s chain *B* (C^α r.m.s.d.s of 1.06 Å for the alignment with 1b6r and 1.07 Å with 3etj), we found that the PO_4^{3-} in 3etj is closer to the ATP γ-phosphate than HCO_3^- and that the SO_4^{2-} in 1b6r is in about the same position as HCO_3^- (Fig. 3*c*). A sulfate or hydrolyzed phosphate

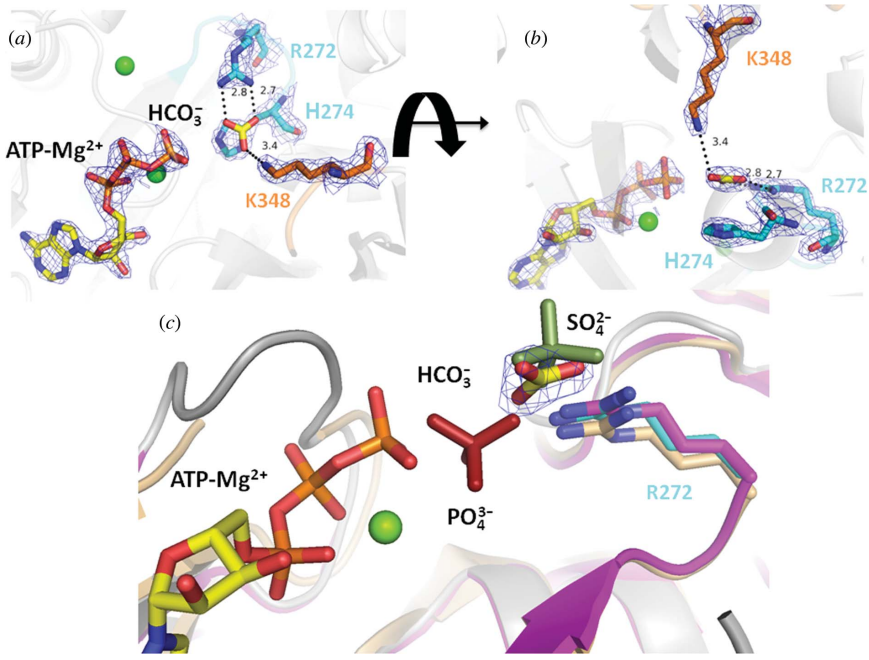

Figure 3
HCO_3^- is shown in the active site of *ba*PurK with ATP and two Mg^{2+} ions (3v4s chain *B*). (*a*) HCO_3^- (red and yellow) interacts with the side chains of Arg272 and His274 (in the J-loop) and Lys348 (in the ABC-loop). The electron-density maps [$2mF_o - DF_c$ map at 2.0σ (0.59 e Å^{-3})] of ATP, Arg272, His274 and Lys348 are shown. The density map for HCO_3^- is not shown here for clarity. (*b*) A rotation of (*a*) to show that HCO_3^-, with its density map, exhibits trigonal planar geometry. (*c*) The previously published PO_4^{3-} group in *E. coli* PurK (dark red; 3etj, chain in light orange) and SO_4^{2-} group (dark green; 1b6r, chain in magenta) are shown in the active site when the structures were overlaid with that of *ba*PurK (3v4s chain *B*, chain in grey and J-loop in cyan). The PO_4^{3-} group is located closer to the γ-phosphate of ATP than HCO_3^- and the SO_4^{2-} group is at about the same location as HCO_3^-. Residue Arg272 corresponds to Arg242 in *E. coli* PurK.

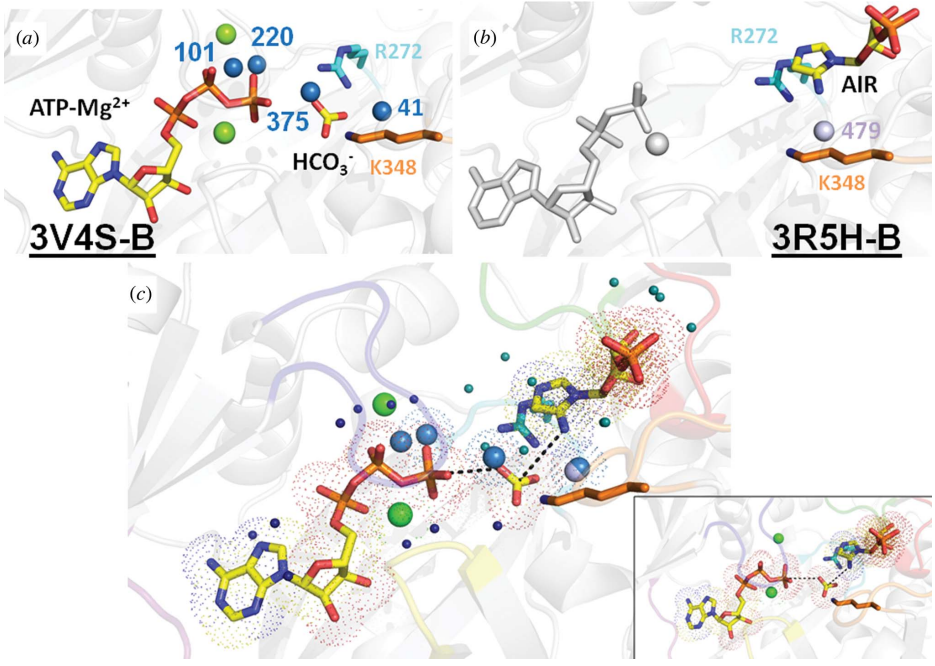

Figure 4
(*a*) ATP, Mg^{2+}, HCO_3^- and key water molecules (waters 41, 101, 220 and 375) in the active site of 3v4s chain *B*. (*b*) AIR and an important water molecule (479, in blue/grey) in the active site of 3r5h chain *B*, used for the model [see text, shown in (*c*)], are shown in color, while ADP-Mg^{2+}, which was not used, is shown in grey. The key residues Arg272 and Lys348 are also shown. (*c*) The model shows HCO_3^- positioned between the γ-phosphate of ATP and the amine (N6) of AIR, with the terminal phosphate P atom 4.5 Å from the O atom in HCO_3^- and the N6 atom in AIR 4.5 Å from the C atom in HCO_3^- (dashed lines). The N6 atom is the site of carboxylation. The side chains of Arg272 and Lys348 are also shown. His274 is not shown for clarity. Important waters in the active site are shown as spheres (those in 3r5h chain *B* in blue/grey and those in 3v4s chain *B* in dark blue). A water (either water 479 in blue/grey from 3r5h chain *B* or water 41 in dark blue from 3v4s chain *B*) is within hydrogen-bonding distance of Lys348, the N6 amine of AIR and HCO_3^-. The van der Waals radii are shown for the reactants with H_2O and without H_2O (bottom inset). Overlapping radii suggest a water-facilitated mechanism for the complete reaction (see Fig. 5). Other water molecules found in the active site are shown as small spheres.

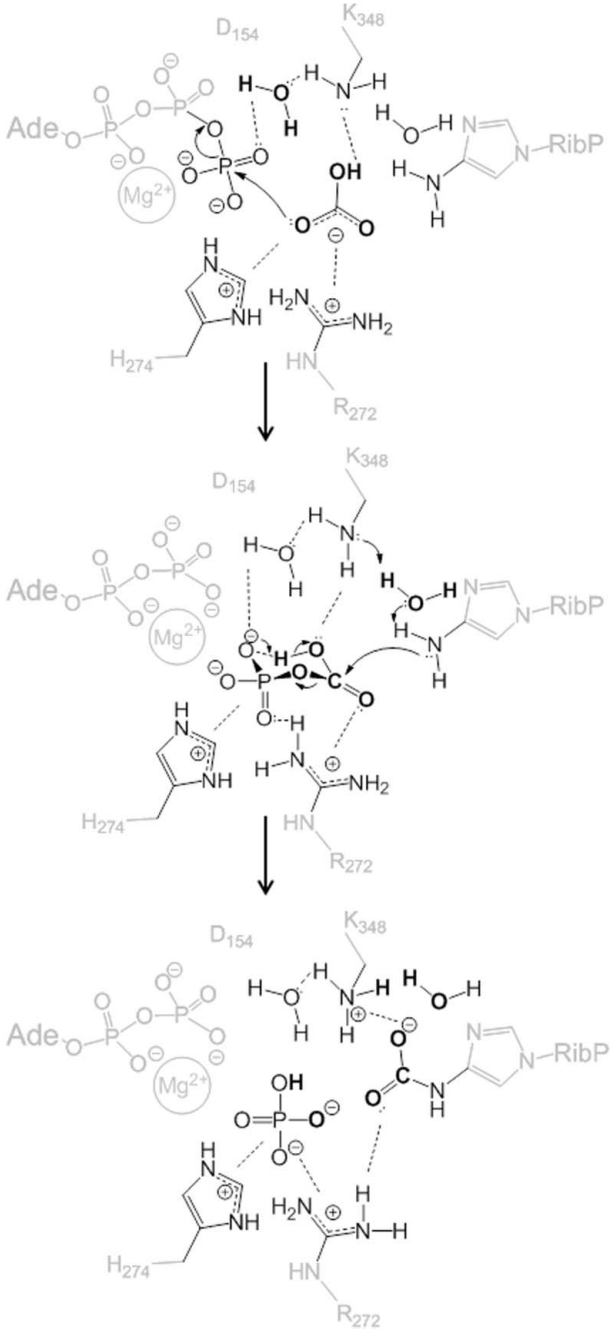

Figure 5
Proposed mechanism for the carboxylation of AIR to N^5-CAIR by *ba*PurK. The binding site characterized in this study shows the involvement of substrates (HCO_3^- and AIR) along with ATP-Mg^{2+}, water molecules and the residues Arg272, His274 and Lys348 in the carboxylation reaction. Lys348 is uncharged owing to nearby hydrophobic side chains (see text) or is deprotonated by Asp154 (top panel). After phosphoryl transfer, which is assisted by a nearby water and/or the bicarbonate (top panel), the bicarbonate is positioned by interacting side chains, allowing nucleophilic attack on the γ-phosphate assisted by the water network (middle panel). The intermediate carboxyphosphate is formed, presumably in a more stable chair conformation (Thoden *et al.*, 2008), and is stabilized by Arg272, His274 and Lys348. The intermediate carboxyphosphate is then shuttled toward the N6 atom (amine) of AIR, which is aided by phosphate repulsion. AIR is primed for carboxylation *via* proton extraction by the nearby water molecule (facilitated by its environment; see Supplementary Fig. S4) in conjunction with Lys348. The carboxyphosphate then breaks down into phosphate and the carboxyl group interacts with AIR to give the product N^5-CAIR (bottom panel).

group shows a protruding tetrahedral shape when placed in the density of the OMIT map.

In the *ba*PurK active site, HCO_3^- is in contact with Arg272 and His274 in the J-loop and with Lys348 in the ABC-loop (Fig. 3). A water molecule (water 41) and the Arg272 side chain both interact with O atoms in HCO_3^- (Fig. 4; Supplementary Table S1), with the two O atoms positioned such that the delocalized negative charge on both O atoms interacts with the delocalized positive charge on the Arg272 side chain to give the guanidinium group and HCO_3^- a coplanar configuration with strong charge–charge interaction and to position the HCO_3^- in a specific location in the active site. In addition to the interaction with the Arg272 side chain and water 41, HCO_3^- also interacts with another water molecule (water 375, which also interacts with the γ-phosphate of ATP and Lys348). His274 and Lys348 both play an important role in stabilizing HCO_3^-. As discussed below, we suggest that both residues facilitate the formation of the product. Notably, all three residues, Arg272, His274 and Lys348, are conserved (Supplementary Fig S1).

4. Discussion

In this study, we find all of the reaction substrates, ATP-Mg^{2+}, HCO_3^-, AIR and a water network, in the active sites of *ba*PurK, but not all together at the same time. The active sites are relatively large, and the structures with different ligands are in general similar except for the B-loop. The B-loop acts as a lid to the active site when ATP is present, as in other ATP-grasp enzymes. The X-ray structures reported in this study allow us to model an active site occupied by all substrates and to propose a reaction mechanism for *ba*PurK.

4.1. Active-site modeling

Since we found (i) ATP-Mg^{2+}, HCO_3^- and AIR in the active site in different structures, (ii) that the backbone structures in each structure, except for a few loop regions, are similar, with r.m.s.d. values for alignment of all C^α atoms of <0.85 Å and (iii) that the ADP part of ATP (3v4s chain *B*) is in approximately the same position as ADP (3r5h chain *B*), we simply removed ADP-Mg^{2+} in 3r5h chain *B*, leaving only AIR in the active site (Fig. 4*b*). We then aligned the two structures to place AIR into the active site of 3v4s chain *B*, which was already occupied by ATP-Mg^{2+}, HCO_3^- and water molecules (Fig. 4*a*; note that the B-loop is in the 'down position' in this structure; see §3.1) to create an active site occupied by all of the substrates (ATP-Mg^{2+}, HCO_3^-, AIR and H_2O; Fig. 4*c*) needed to form the product N^5-CAIR. In this model structure, we find that HCO_3^- is equidistant from ATP and AIR, with the P atom of the terminal phosphate at a distance of 4.5 Å from an O atom in HCO_3^- and the N6 atom of AIR at a distance of 4.5 Å from the C atom in HCO_3^-. The N6 atom is the exocyclic N atom (Fig. 2, inset) and is the site of carboxylation to yield N^5-CAIR. There are also a number of water molecules between the γ-phosphate (two within 4.0 Å), HCO_3^- and AIR. The positive charges on the side chains of Arg272, His274 and

Lys348 are all about 3.0 Å from the O atoms of HCO_3^- (Supplementary Table S1).

4.2. Proposed mechanism

PurK catalyzes the reaction of AIR and HCO_3^- to form N^5-CAIR in the presence of ATP-Mg^{2+} in the *de novo* purine-biosynthesis pathway. We identify three important residues in the active site: Arg272, His274 and Lys348. With the precise position and orientation of HCO_3^- in *ba*PurK now determined, we propose a structure in which its active site is filled with the substrates AIR and HCO_3^- as well as with ATP-Mg^{2+} and a water network. With this structure, we propose the reaction mechanism shown in Fig. 5.

4.2.1. ATP hydrolysis. It has been suggested that ATP is hydrolyzed and the phosphate is transferred to HCO_3^- to form an intermediate carboxyphosphate (Mueller *et al.*, 1994). Two general mechanisms for ATP hydrolysis and phosphoryl transfer have been proposed: the 'associative' and the 'dissociative' mechanisms (Harrison & Schulten, 2012). In the associative mechanism, an O atom that is within 4.0 Å, such as the oxygen in a 'lytic' water (O_w), interacts with the γ-phosphate, eventually leading to a transition state of O_w associated with the γ-phosphate. In the dissociative mechanism, water is hydrogen-bonded to a terminal O atom of the γ-phosphate to facilitate the breakdown of the $P^{\beta}-O-P^{\gamma}$ 'bridge' by lengthening the $O-P^{\gamma}$ bond required for the dissociative mechanism (Harrison & Schulten, 2012). In *ba*PurK, waters 101 and 375 are 3.1 and 3.6 Å away from the γ-phosphate, respectively. Thus, it is possible to follow the associative mechanism with O_{w101} or O_{w375} interacting with the γ-phosphate that eventually breaks down to generate ADP and $H_2PO_4^-$ (Fig. 5, top). $H_2PO_4^-$ will subsequently react with HCO_3^- to form carboxyphosphate and water (Fig. 5, center). On the other hand, at a distance of only 4.5 Å from the γ-phosphate, the O atom of HCO_3^- may still associate with the γ-phosphate directly, leading to the formation of the carboxyphosphate intermediate (Fig. 5, top and center). It is also possible to follow the dissociative mechanism suggested by Harrison & Schulten (2012). Waters 101, 220 and 375 are all within hydrogen-bonding distance of a terminal O atom in the γ-phosphate, which can facilitate the breakdown of the $P^{\beta}-O-P^{\gamma}$ bridge by lengthening the $O-P^{\gamma}$ bond to give ADP and $H_2PO_4^-$ (Fig. 5, top). $H_2PO_4^-$ will subsequently react with HCO_3^- to lose a water to form carboxyphosphate. In short, with a water network close to the γ-phosphate and a nearby HCO_3^-, the structural information does not allow us to distinguish one pathway over another for ATP to lose its terminal phosphate, leading to the proposed carboxyphosphate intermediate, in *ba*PurK. All of the mentioned pathways are possible.

4.2.2. Carboxyphosphate intermediate and reaction mechanism. Carboxyphosphate is short-lived (Mueller *et al.*, 1994), with a half-life of 70 ms in solution (Thoden *et al.*, 1997). Its half-life in *E. coli* carbamoyl phosphate synthetase in 67% DMSO was estimated to be 2.5 min (Powers & Meister, 1978). This chemical instability explains its lack of characterization as an intermediate in the purine-biosynthetic pathway (Mueller *et al.*, 1994) or its capture in an enzyme structure. In *ba*PurK, we propose that the carboxyphosphate intermediate is formed when the van der Waals radii of the atoms in γ-phosphate, water and HCO_3^- (Fig. 4*c*) overlap slightly (Figs. 4*c* and 5). HCO_3^- is positioned in a favorable position for reaction by the side chains of Arg272 and His274 in the J-loop and of Lys348 in the ABC-loop (Fig. 5, left). Next to the carboxyphosphate intermediate is the amine (protonated N6 atom) of AIR. In order for the carboxyphosphate to carboxylate AIR to form N^5-CAIR, the amine needs to be deprotonated. In *ac*PurK, the side chain of *Tyr152* (equivalent to Tyr153 in *ba*PurK, where the italic font indicates that this is the numbering system for *ac*PurK) interacts with the imidazole ring of AIR, and *Asp153* acts as a base to deprotonate the amine (see Fig. 2 in Dewal & Firestine, 2013; Thoden *et al.*, 2010). However, in *ba*PurK the distances between the closest C atoms in Tyr153 and the imidazole ring and between both O atoms of Asp154 and the N6 atom of AIR are larger than 6 Å. Thus, in *ba*PurK no residues are nearby to carry out the deprotonation of AIR. We suggest a proton-relay mechanism, with Lys348 being uncharged (Fig. 5, top) since its pK_a is lowered by hydrophobic side chains (Ho *et al.*, 2009) nearby (within 7 Å), namely Ile217, Leu218 and Leu316. It is notable that these hydrophobic side chains are mostly conserved in the 12 species mentioned before. It is also possible that Asp154, with the O atom 4.2 Å from the N atom of Lys348, contributes to the uncharged state of Lys348. With the uncharged Lys348 and its nearby water molecules, the Lys348–water 'complex' acts as a base to deprotonate the amine of AIR (Fig. 5, center). The carboxyphosphate interacts with the deprotonated AIR to give the product N^5-CAIR (Fig. 5, bottom). It should be noted that the product N^5-CAIR is not stable. At 303 K and pH 7.8 it exhibits a half-life of 0.9 min and is even less stable at lower pH (Mueller *et al.*, 1994). Notably important in our proposed mechanism is that HCO_3^- is in contact with the conserved residues Arg272, His274 and Lys348 and with water molecules. In this mechanism, the specific side-chain geometry, and not the charge, of Arg272 is important in promoting product formation, and this explains why the *R271K* mutation in *ac*PurK leads to a less active enzyme (Dewal & Firestine, 2013). Since the residues involved in the proposed mechanism, Arg272, His274 and Lys348, are conserved residues and since the active sites of different prokaryotic PurK are similar, we suggest that our proposed mechanism may also apply to other prokaryotic PurK systems.

5. Conclusion

The structures presented here provide further insights into the reaction mechanism of *ba*PurK by showing bicarbonate bound in the active site near the conserved residues Arg272, His274 and Lys348 and highlighting the importance of the active-site water network. We suggest a mechanism facilitated by water molecules for the formation of the intermediate and the product.

This work was supported in part by the Defense Threat Reduction Agency (DTRA) program, contract HDTRA1-11-C-0011. Data were collected SER-CAT, at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109- Eng-38. The authors thank Loredona Huma in the MEJ laboratory for preparing AIR. We would like to thank the reviewers and the coeditor. Their insightful comments and the patience of the coeditor in allowing multiple revisions have greatly improved this manuscript.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Brugarolas, P., Duguid, E. M., Zhang, W., Poor, C. B. & He, C. (2011). *Acta Cryst.* D**67**, 707–715.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Dewal, M. B. & Firestine, S. M. (2013). *Biochemistry*, **52**, 6559–6567.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Fawaz, M. V., Topper, M. E. & Firestine, S. M. (2011). *Bioorg. Chem.* **39**, 185–191.

Firestine, S. M., Misialek, S., Toffaletti, D. L., Klem, T. J., Perfect, J. R. & Davisson, V. J. (1998). *Arch. Biochem. Biophys.* **351**, 123–134.

Firestine, S. M., Poon, S.-W., Mueller, E. J., Stubbe, J. & Davisson, V. J. (1994). *Biochemistry*, **33**, 11927–11934.

Harrison, C. B. & Schulten, K. (2012). *J. Chem. Theory Comput.* **8**, 2328–2335.

Ho, M.-C., Ménétret, J.-F., Tsuruta, H. & Allen, K. A. (2009). *Nature (London)*, **459**, 393–397.

Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kappock, T. J., Ealick, S. E. & Stubbe, J. (2000). *Curr. Opin. Chem. Biol.* **4**, 567–572.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.

Marie, S., Heron, B., Bitoun, P., Timmerman, T., Van Den Berghe, G. & Vincent, M. F. (2004). *Am. J. Hum. Genet.* **74**, 1276–1281.

Meyer, E., Kappock, T. J., Osuji, C. & Stubbe, J. (1999). *Biochemistry*, **38**, 3012–3018.

Mueller, E. J., Meyer, E., Rudolph, J., Davisson, V. J. & Stubbe, J. (1994). *Biochemistry*, **33**, 2269–2278.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Novotny, M., Seibert, M. & Kleywegt, G. J. (2007). *Acta Cryst.* D**63**, 270–274.

Patey, C. A. H. & Shaw, G. (1973). *Biochem. J.* **135**, 543–545.

Powers, S. G. & Meister, A. (1978). *J. Biol. Chem.* **253**, 1258–1265.

Sivendran, S., Patterson, D., Spiegel, E., McGown, I., Cowley, D. & Colman, R. F. (2004). *J. Biol. Chem.* **279**, 53789–53797.

Thoden, J. B., Holden, H. M. & Firestine, S. M. (2008). *Biochemistry*, **47**, 13346–13353.

Thoden, J. B., Holden, H. M., Paritala, H. & Firestine, S. M. (2010). *Biochemistry*, **49**, 752–760.

Thoden, J. B., Holden, H. M., Wesenberg, G., Raushel, F. M. & Rayment, I. (1997). *Biochemistry*, **36**, 6305–6316.

Thoden, J. B., Kappock, T. J., Stubbe, J. & Holden, H. M. (1999). *Biochemistry*, **38**, 15480–15492.

Thoden, J. B., Wesenberg, G., Raushel, F. M. & Holden, H. M. (1999). *Biochemistry*, **38**, 2347–2357.

Tuntland, M. L., Johnson, M. E., Fung, L. W.-M. & Santarsiero, B. D. (2011). *Acta Cryst.* D**67**, 870–874.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.

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

Ye, D., Wei, M., McGuire, M., Huang, K., Kapadia, G., Herzberg, O, Martin, B. M. & Dunaway-Mariano, D. (2001). *J. Biol. Chem.* **276**, 37630–37639.

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