Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Expression, purification, crystallization and preliminary X-ray diffraction crystallographic study of PurH from *Escherichia coli*

Xiaoting Qiu,[a,b] Ye Yuan[a,b] and Yongxiang Gao[a,b]*

[a]Hefei National Laboratory for Physical Sciences at Microscale, School of Life Sciences, University of Science and Technology of China, 96 Jinzhai Road, Hefei, Anhui 230026, People's Republic of China, and [b]Key Laboratory of Structural Biology, Chinese Academy of Sciences, 96 Jinzhai Road, Hefei, Anhui 230026, People's Republic of China

Correspondence e-mail: yxgao@ustc.edu.cn

Received 28 August 2011
Accepted 28 September 2011

In bacteria and eukaryotes, the last two steps of *de novo* purine biosynthesis are catalyzed by bifunctional purine-biosynthesis protein (PurH), which is composed of two functionally independent domains linked by a flexible region. The N-terminal domain possesses IMP cyclohydrolase activity and the C-terminal domain possesses aminoimidazole-4-carboxamide ribonucleotide transformylase activity. This study reports the expression, purification, crystallization and preliminary X-ray crystallographic analysis of PurH from *Escherichia coli* with an N-terminal His_6 tag. The crystals diffracted to a maximum resolution of 3.05 Å and belonged to the monoclinic space group $P2_1$, with unit-cell parameters a = 76.37, b = 132.15, c = 82.64 Å, β = 111.86°.

1. Introduction

The production of inosine 5′-monophosphate (IMP) from small-molecule precursors in organisms is known as *de novo* purine biosynthesis (Buchanan & Hartman, 1959) and is composed of ten enzymatic catalytic steps in higher organisms, while an additional enzyme is required for the sixth step in prokaryotes. IMP is then converted to guanosine 5′-monophosphate (GMP) or adenosine 5′-monophosphate (AMP) by subsequent enzymes. IMP, AMP and GMP are also generated *via* the purine-salvage pathway, which is the sole pathway for obtaining purine nucleotides in some parasitic organisms (Zhang *et al.*, 2008). Because of the critical role of purine nucleotides in the synthesis of RNA and DNA, the *de novo* purine-biosynthetic pathway is considered to be an important target for anticancer, antiviral and antibacterial drug design.

The last two steps in the purine-biosynthetic pathway are the conversion of aminoimidazole-4-carboxamide ribonucleotide (AICAR) to the final product IMP (Fig. 1). In bacteria and eukaryotes these two steps are catalyzed by the bifunctional enzyme AICAR transformylase (AICAR Tfase)/IMP cyclohydrolase (IMPCH) (EC 2.1.2.3), also known as bifunctional purine-biosynthesis protein (PurH). This enzyme has become a target for the development of anticancer therapeutics, especially for the investigation of specific antifolate reagents (Cheong *et al.*, 2004; Wolan *et al.*, 2003) and nonfolate inhibitors (Li *et al.*, 2004; Xu *et al.*, 2004), which are analogues of cofactor N^{10}-formyltetrahydrofolate (10-f-THF) which can completely inhibit AICAR Tfase activity.

Bifunctional purine-biosynthesis protein from *Escherichia coli* (EcPurH) is encoded by the *purH* gene. This enzyme is composed of two domains linked by a flexible region. The N-terminal domain possesses IMPCH activity and the C-terminal domain possesses AICAR Tfase activity. Coupling of the two domains has been shown to be essential for the catalytic process, as the AICAR Tfase reaction

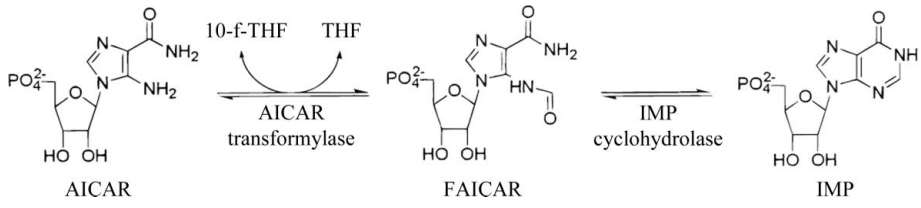


Figure 1
The reactions catalyzed by PurH (Wall *et al.*, 2000).

favours the reverse direction by itself and the irreversible cyclization of 5-formyl-AICAR (FAICAR) to IMP drives formyl transfer in the forward direction (Xu *et al.*, 2007).

The structures of several prokaryotic and eukaryotic PurH enzymes have been determined (Axelrod *et al.*, 2008; Cheong *et al.*, 2004; Greasley *et al.*, 2001; Wolan *et al.*, 2002, 2003; Xu *et al.*, 2004, 2007), which reveal that the relative positions of the two domains in PurH differ in different species. EcPurH has inserted sequences in its N-terminal domain, a deleted sequence in its interdomain region and deleted or inserted sequences in its C-terminal domain compared with the sequences of PurH from other species for which structures have been solved (Fig. 2). These sequence variations could cause the

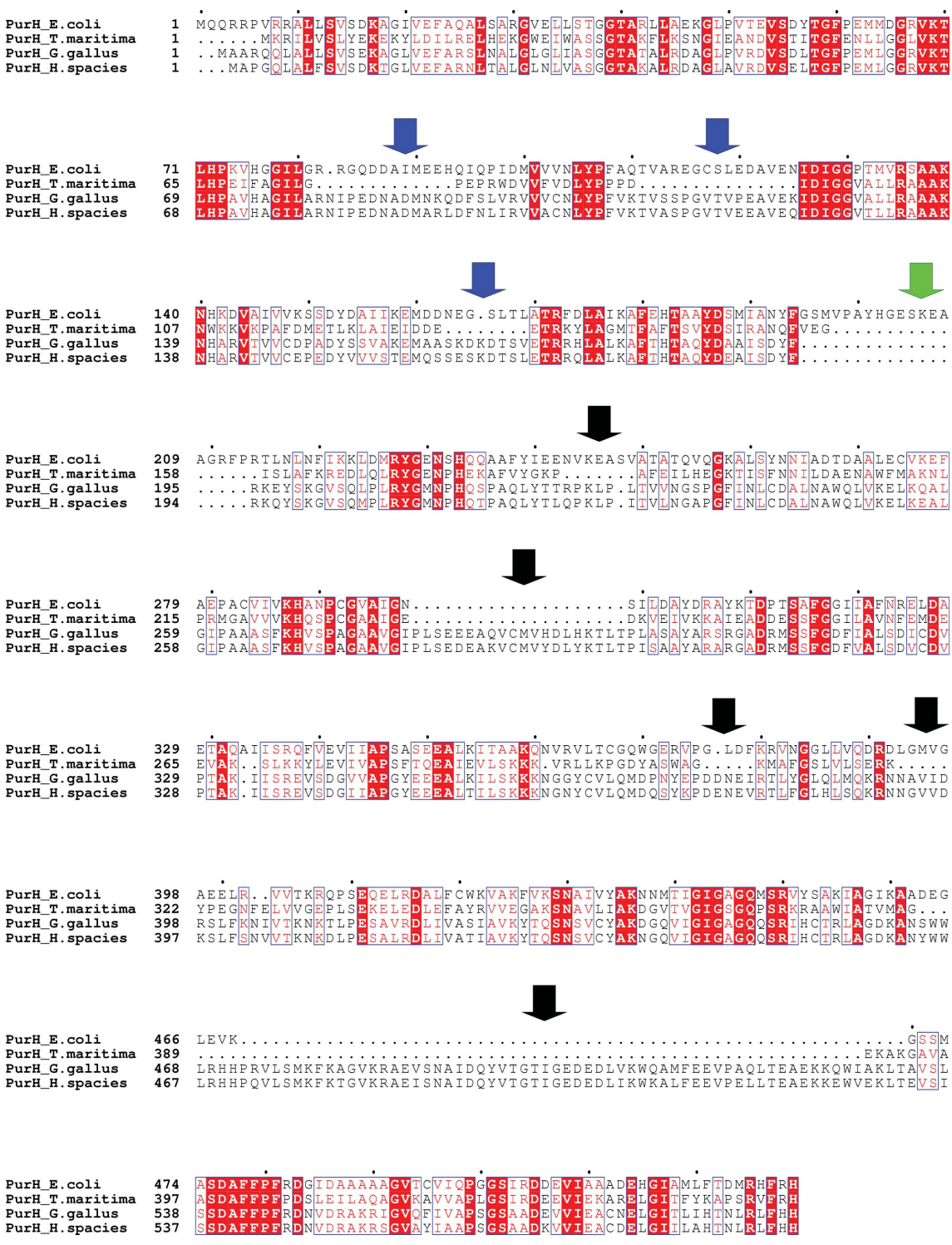


Figure 2
Multi-sequence alignment of EcPurH with PurH enzymes from other species for which structures have been solved. The inserted sequences in the N-terminal domain, the deleted sequence in the interdomain region and the deleted and inserted sequences in the C-terminal domain of EcPurH are indicated by blue, green and black arrows respectively. The alignment was generated using the *ClustalW* web server.

interdomain positioning in EcPurH to differ from that in PurH from other species, which could make substrate transfer between the two domains of EcPurH different from that in other PurH enzymes. To explore the spatial relationship of the two catalytic active sites in EcPurH and its impact on its catalytic rate, the expression, purification, crystallization and preliminary X-ray crystallographic analysis of this enzyme are reported in this study.

2. Materials and methods

2.1. Cloning, expression and purification

The cDNA of full-length EcPurH was obtained *via* PCR from the *E. coli* strain K12 genome and was cloned into pET28a (Novagen) excised using *Nde*I and *Xho*I to create recombinant EcPurH with an N-terminal hexahistidine tag (MGSSHHHHHHLVPRGSH). The sequence of the cDNA was confirmed by DNA sequencing. A single colony of *E. coli* Rosetta (DE3) (Novagen) bacteria harbouring the expression vector was cultured in 8 ml Luria–Bertani broth overnight and was then used to inoculate 0.8 l medium containing 50 µg ml^{-1} kanamycin. The cells were grown at 310 K for 2.5 h until the $OD_{600\,nm}$ reached 0.5–0.8 and protein expression was then induced for 24 h with 0.25 m*M* isopropyl *β*-D-1-thiogalactopyranoside (IPTG) at 289 K. The bacteria were collected and resuspended in 50 ml binding buffer (20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl). After disrupting the cells by sonication, the bacteria were centrifuged at 15 200*g* for 0.5 h. The clean lysate supernatant was loaded onto Ni–NTA agarose (GE Healthcare) resin pre-equilibrated with binding buffer. The tagged protein was eluted with 30 ml binding buffer containing 500 m*M* imidazole, which was then concentrated for further purification using Superdex 200 gel-filtration (GE Healthcare) chromatography eluted with binding buffer containing 5 m*M* dithiothreitol (DTT). The retention volume corresponding to the target protein indicated that it was a monomer in solution. The fractions containing the peak were pooled, exchanged with buffer *A* (20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 5 m*M* DTT) and then further purified using Q-Sepharose Fast Flow (GE Healthcare) chromatography eluted with a linear gradient of NaCl from 0.1 to 0.5 *M*. The fractions from the peak corresponding to the target protein (Fig. 3*a*) were pooled and concentrated to 56 mg ml^{-1} using a 10 kDa cutoff Amicon centrifugal ultrafilter concentrator (Millipore). Examination of the purified protein by SDS–PAGE revealed a single band corresponding to the expected molecular weight (Fig. 3*b*). The protein concentration was measured using the BCA Protein Assay Kit (Pierce).

2.2. Lysine methylation

EcPurH contains a relatively large amount of lysine (28 lysines in 592 residues), which could prevent crystallization. Therefore, lysine methylation was performed basically as described previously (Walter *et al.*, 2006) to improve the quality of crystallization. The purified EcPurH was diluted to less then 1 mg ml^{-1} in 50 m*M* HEPES pH 7.5, 250 m*M* NaCl. 20 µl freshly prepared 1 *M* dimethylamine–borane complex (ABC; Fluka) and 40 µl 1 *M* formaldehyde (Fluka) were then added per millilitre of protein solution. The reaction was carried out at 277 K. After 2 h, a further 20 µl 1 *M* ABC and 40 µl 1 *M* formaldehyde were added per millilitre of solution and the mixture was incubated for a further 2 h. 10 µl 1 *M* ABC per millilitre of solution was then added and the mixture was incubated at 277 K overnight. Finally, the reaction solution was concentrated and applied onto a Superdex 200 gel-filtration chromatography column pre-equilibrated with buffer *A* in order to remove ABC and formaldehyde.

2.3. Crystallization

Preliminary screening for initial crystallization conditions for EcPurH without reductive lysine methylation was performed by the sitting-drop vapour-diffusion method using ProPlex (Molecular Dimensions) at 287 K by mixing 1 µl 56 mg ml^{-1} protein solution with an equal volume of reservoir solution in 48-well plates. Small block-shaped crystals were obtained from the condition 0.1 *M* sodium acetate pH 5.0, 1 *M* ammonium sulfate. The conditions were further optimized using various concentrations of ammonium sulfate *versus* a pH range of 4.5–5.5. The diffraction quality of the crystals from the optimal conditions (Fig. 4*a*) was poor. Microseeding of these crystals did not improve the diffraction quality. Initial crystallization conditions of EcPurH with reductive lysine methylation were screened using the same kit and using the same method as used for EcPurH without reductive lysine methylation. Clusters of plate-shaped crystals (Fig. 4*b*) appeared in the condition 0.8 *M* sodium/potassium hydrogen phosphate pH 7.5 after one week and reached maximum size after one month; these crystals were used for data collection.

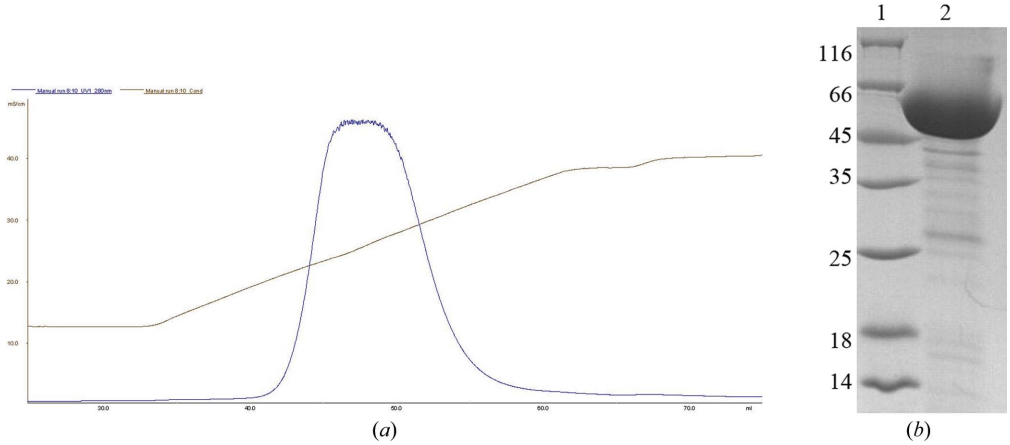
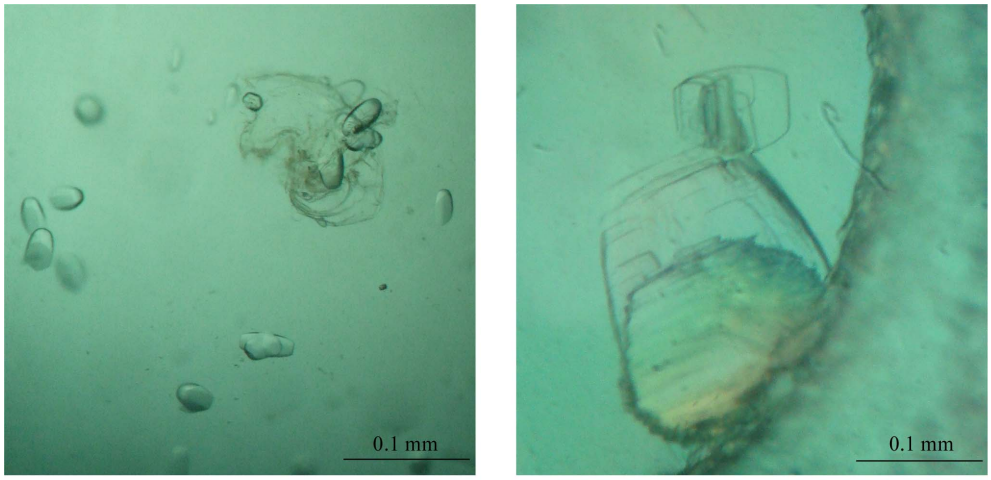


Figure 3
(*a*) The peak corresponding to EcPurH in the ion-exchange chromatogram of EcPurH purified using Q-Sepharose Fast Flow gel-filtration chromatography eluted with a linear gradient of NaCl from 0.1 to 0.5 *M*. The *y* axis and the *x* axis represent the conductivity per centimetre and the retention volume, respectively. (*b*) SDS–PAGE of EcPurH. The protein was analyzed on 12% SDS–PAGE and stained with Coomassie Blue. Lane 1, molecular-weight markers (labelled in kDa); lane 2, EcPurH after three purification steps.



(*a*) (*b*)

Figure 4
(*a*) Crystals of EcPurH without reductive lysine methylation as grown in 0.1 *M* sodium acetate pH 5.0, 1 *M* ammonium sulfate. (*b*) Crystals of EcPurH after reductive lysine methylation as grown in 0.8 *M* sodium/potassium hydrogen phosphate pH 7.5 and used in the diffraction experiment.

Table 1
Data-collection statistics for EcPurH.

Values in parentheses are for the highest resolution shell.

Space group	$P2_1$
Wavelength (Å)	0.9792
Unit-cell parameters (Å, °)	a = 76.37, b = 132.15, c = 82.64, β = 111.86
Resolution (Å)	46.57–3.05 (3.21–3.05)
Total No. of observations	110533 (14731)
No. of unique reflections	28296 (3977)
Multiplicity	4.0 (3.7)
Completeness (%)	97.5 (94.7)
Average $I/\sigma(I)$	8.0 (2.1)
$R_{\rm merge}$†	0.140 (0.655)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity of a reflection and $\langle I(hkl)\rangle$ is the mean intensity of reflection *hkl*.

2.4. X-ray diffraction data collection and processing

Before data collection, one piece of the crystal cluster was picked up and quick-soaked in a cryoprotectant solution consisting of 25%(*v*/*v*) glycerol, 0.8 *M* sodium/potassium hydrogen phosphate pH 7.5 and flash-cooled in a nitrogen stream at 100 K. X-ray diffraction data were collected on beamline 17U1 of the Shanghai Synchrotron Radiation Facility (SSRF) using an MX-225 CCD detector (MAR Research). The crystal-to-detector distance was kept at 240 mm and the crystal was rotated through a total of 200°, with 1° rotation and an exposure time of 1.2 s per frame. The crystal diffracted to a maximum resolution of 3.05 Å. All data were indexed and integrated with *iMOSFLM* (Battye *et al.*, 2011) and scaled using *SCALA* (Evans, 2006) from the *CCP4* suite (Winn *et al.*, 2011). Analysis of the systematic absences in the diffraction data characterized the space group as $P2_1$. The final statistics of data collection and processing are listed in Table 1.

3. Results and discussion

Full-length PurH from *E. coli* strain K12 was expressed and purified to homogeneity with an N-terminal hexahistidine tag. Because of the poor diffraction quality of the crystals of EcPurH, lysine methylation was performed. EcPurH with reductive lysine methylation crystallized from 0.8 *M* sodium/potassium hydrogen phosphate pH 7.5. The crystals diffracted to a maximum resolution of 3.05 Å and belonged to the monoclinic space group $P2_1$, with unit-cell parameters a = 76.37, b = 132.15, c = 82.64 Å, β = 111.86°. The resulting $R_{\rm merge}$ was 14.0% overall and the average mosaicity was 1.02°. The highest probability of three molecules per asymmetric unit suggested a Matthews coefficient of 2.18 Å^3 Da^{-1} (Matthews, 1968), with a solvent content calculated as 43.62%.

To accommodate the possible difference in the relative positioning of the two domains of EcPurH compared with homologous structures, structure determination will be attempted *via* the molecular-replacement method using the structures of the N-terminal domain of PurH from *Gallus gallus* (PDB entry 1m9n; Wolan *et al.*, 2002) and the C-terminal domain of PurH from *Thermotoga maritima* (PDB entry 1wzn; Axelrod *et al.*, 2008) as independent entities in the search model as these two domains share the highest sequence similarity with the corresponding domains of EcPurH among the PurH structures solved.

Financial support for this project was provided by research grants from the Junior Scientist Funds of USTC (grant No. KA207000007), the Chinese National Natural Science Foundation (grant Nos. 30025012 and 10979039) and the Chinese Ministry of Science and Technology (grant Nos. 2006CB806500, 2006CB910200 and 2006AA02A318).

References

Axelrod, H. L. *et al.* (2008). *Proteins*, **71**, 1042–1049.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Buchanan, J. M. & Hartman, S. C. (1959). *Adv. Enzymol. Relat. Areas Mol. Biol.* **21**, 199–261.
Cheong, C.-G., Wolan, D. W., Greasley, S. E., Horton, P. A., Beardsley, G. P. & Wilson, I. A. (2004). *J. Biol. Chem.* **279**, 18034–18045.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Greasley, S. E., Horton, P., Ramcharan, J., Beardsley, G. P., Benkovic, S. J. & Wilson, I. A. (2001). *Nature Struct. Biol.* **8**, 402–406.
Li, C., Xu, L., Wolan, D. W., Wilson, I. A. & Olson, A. J. (2004). *J. Med. Chem.* **47**, 6681–6690.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Wall, M., Shim, J. H. & Benkovic, S. J. (2000). *Biochemistry*, **39**, 11303–11311.
Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I. & Grimes, J. M. (2006). *Structure*, **14**, 1617–1622.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Wolan, D. W., Greasley, S. E., Beardsley, G. P. & Wilson, I. A. (2002). *Biochemistry*, **41**, 15505–15513.
Wolan, D. W., Greasley, S. E., Wall, M. J., Benkovic, S. J. & Wilson, I. A. (2003). *Biochemistry*, **42**, 10904–10914.
Xu, L., Chong, Y., Hwang, I., D'Onofrio, A., Amore, K., Beardsley, G. P., Li, C., Olson, A. J., Boger, D. L. & Wilson, I. A. (2007). *J. Biol. Chem.* **282**, 13033–13046.
Xu, L., Li, C., Olson, A. J. & Wilson, I. A. (2004). *J. Biol. Chem.* **279**, 50555–50565.
Zhang, Y., Morar, M. & Ealick, S. E. (2008). *Cell. Mol. Life Sci.* **65**, 3699–3724.