PROTEINS: Structure, Function, and Bioinformatics 61:674–676 (2005)

STRUCTURE NOTE

Crystal Structure of PurE (BA0288) from *Bacillus anthracis* at 1.8 Å Resolution

Mark P. Boyle, Anne K. Kalliomaa, Vladimir Levdikov, Elena Blagova, Mark J. Fogg,* James A. Brannigan, Keith S. Wilson, and Anthony J. Wilkinson

Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, United Kingdom

Introduction. As part of a structural genomics program, Structural Proteomics in Europe (SPINE), we are determining protein structures from the causative agent of anthrax, *Bacillus anthracis*, a Gram-positive spore-forming bacterium. Among initial candidates for crystallographic analysis are enzymes involved in nucleotide biosynthesis. The BA0288 gene (www.tigr.org) of *B. anthracis* encodes a protein with 57% amino acid sequence identity to the *Escherichia coli* 5′-phosphoribosyl-5-aminoimidazole carboxylase (PurE).[1] PurE proteins are highly conserved and are designated as Class I or Class II according to their enzymatic activity. Class I enzymes, found in yeast, plants, and prokaryotes, catalyze the second of a two-step conversion of 5-aminoimidazole ribonucleotide (AIR), via the intermediate N^5-carboxyaminoimidazole ribonucleotide (N^5-CAIR), to 4-carboxy-5-aminoimidazole ribonucleotide (CAIR).[2,3] The conversion of AIR to N^5-CAIR is catalyzed by N^5-CAIR synthetase (PurK) in the presence of ATP and bicarbonate. Class II enzymes from higher eukaryotes catalyze the conversion of AIR to CAIR directly, in the presence of bicarbonate or CO_2.[4] Therefore, Class I PurE proteins function as phosphoribosylaminoimidazole mutases, while Class II enzymes are carboxylases. This difference in activity could be exploited to provide potential targets for antibacterial therapies.[5]

Results and Discussion. The structure of PurE was determined to 1.8 Å resolution by molecular replacement using the coordinate set for the *Escherichia coli* orthologue (PDB code 1QCZ[6]) as a search model. Data collection, refinement and model-building statistics are summarized in Table I. The refined model consists of eight protein molecules (residues −7–161) for chains A–H and a total of 1219 water molecules. Residues 156–162 are not clearly defined in the electron-density maps for most subunits and are assumed to be disordered, although complete backbone chains can be traced for subunits B, E, and F. The Matthews' coefficient (V_m) for the crystals is 2.4 Å^3/Da, and the estimated solvent content is 48.5%. The Ramachandran plot produced by PROCHECK 3.4[7] shows that 93.9% of residues are in the most favored regions with 6.1% in additional allowed regions.

All the observed X-ray data in the resolution range 50.0–1.8 Å were used in the structure solution and refinement. While the R_{sym} value is high (60%) in the outer resolution shell, the value of $I/\sigma(I)$ is 2.5, indicating these data are significant. The maximum likelihood program REFMAC5,[8] which we used for refinement, is designed to handle such weak data in a robust manner and accord them appropriate weights. This is confirmed by the value of $R_{free} = 0.288$ in the outer shell, which should be compared to a theoretical R-factor value for a random atom model of 0.586. For this *B. anthracis* PurE structure, the inclusion of all data in the refinement and map calculation considerably aided the building of the rather poorly ordered N-terminal region of two of the eight chains (C and G). The residues (SHHHHHHMKS) in these regions correspond to seven from the purification tag plus three from PurE itself.

The *B. anthracis* PurE monomer [Fig. 1(a)] consists of a five-stranded, parallel β-sheet with a $\beta_2\beta_1\beta_3\beta_4\beta_5$ strand topology together with six α-helices in a 'flavodoxin-like' fold.[9] The strands and helices α_1–α_5 form a compact $(\beta\alpha)_5$ globular domain with helices α_1 and α_5 packing against one face of the β-sheet and helices α_2, α_3, and α_4 packing onto the opposite face. Helix α_6 protrudes away from the monomer, as illustrated in Figure 1(a). The PurE octamer is constructed as a dimer of tetramers, which are related by a two-fold symmetry axis perpendicular to the four-fold symmetry axis. Viewed down the four-fold axis, the assembly has the shape of a square from which the corners have been cut [Fig. 1(b)]. The protruding helix α_6 makes numerous contacts with helix α_2 and strand β_2 of the neighboring subunit, thus contributing extensively to the contacts between adjacent monomers in the tetramer. The total buried surface area in the octamer is over 25,400 Å^2.

Co-crystallisation experiments identified four active-site clefts on each of the four-fold symmetric surfaces of the octamer in *E. coli*.[6] The binding of sulfate ions in the same regions identified conserved residue positions in the PurE crystal structure from *Thermotoga maritima* (PDB code: 1O4V)[10] [Fig. 1(c)]. The analogous putative active site of *B. anthracis* PurE is located at the interface of three subunits.

Grant sponsor: European Commission as SPINE; Grant number: QLG2-CT-2002-00988. Grant sponsor: Wellcome Trust (to J.A.B.). Grant sponsor: BBSRC (to M.P.B. and V.M.L.)

*Correspondence to: Mark J. Fogg, Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, UK. E-mail: fogg@ysbl.york.ac.uk

Received 31 January 2005; Accepted 28 February 2005

Published online 1 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20599

TABLE I. Data Collection and Refinement Parameters

Space group	*C*2
Unit cell parameters (Å)	***a*** = 168.2, ***b*** = 76.4, ***c*** = 102.6, β = 96.7°
Wavelength (Å)	0.8700
Resolution range (Å)	50.0–1.80 (1.85–1.80)
Unique reflections	113681
Completeness (%)	99.5 (95.4)
R_{merge} (%)	8.1 (60)
Mean $I/\sigma(I)$	17.5 (2.5)
Redundancy	3.7 (3.3)
R_{factor} [a] (%)	16.9 (23.4)
R_{free} (%)	20.5 (28.8)
RMSD bond length (Å)	0.012
RMSD bond angle (°)	1.29
Average isotropic *B*-value (Å^2)	16.5

[a] $R_{factor} = \Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{free} is R_{factor} calculated with 5.1% of the reflections chosen at random and omitted from refinement. Data in brackets correspond to the highest resolution shell.

The first subunit contributes important 'P-loop' conserved residues Gly9, Ser10, and Asp12, together with residues 38–42 (SAHRT) and 69–73 (AHLPG), from loops β_2-α_2 and β_3-α_3, described as the 'forties' and 'seventies' loops, respectively, in *E. coli*.[6] These residues form one wall of the active site. The second adjacent subunit forms the opposite wall and contributes three conserved residues (Gln104, Met105, and Pro106) from the end of helix α_4. These motifs are fully conserved in all three PurE structures [Fig. 1(d)] and are closely superimposable. The root mean square deviation (RMSD) in C_α positions following least squares structural superposition of residues 1–155 of the *E. coli* and *B. anthracis* PurE coordinates is 0.61 Å[11] with no large deviations from the mean. This compares with a RMSD of 0.51 Å in the positions of 148 equivalent atoms between the PurE proteins from *B. anthracis* and *Thermotoga maritima*, which are 60.3% identical in sequence.

The Class I PurE-catalyzed reaction is the sixth and only carbon–carbon bond-forming step[2,3] on the pathway for *de novo* synthesis of inosine monophosphate, which is the first complete nucleotide formed during purine biosynthesis.[1,12] Here we report the first crystal structure of a PurE from a Gram-positive organism, revealing an octameric arrangement of subunits. Analysis of the *B. anthracis* PurE structure will provide further information on this important pathway for the bacterium's viability.

Materials and Methods. *Protein production and crystallization.* The coding sequence of BA0288 was amplified by polymerase chain reaction (PCR) from *B. anthracis* genomic DNA using KOD Hot-Start DNA polymerase (Novagen) and complementary gene-specific primers to which were appended sequences to facilitate ligation-independent cloning (LIC).[13] For LIC, the PCR amplification product was treated with T4 DNA polymerase in the presence of dATP to generate 5′ single stranded overhangs at both ends of the fragment, through the enzyme's combined 3′-5′ exonuclease and DNA polymerase activities. Complementary 5′ single-stranded overhangs were generated in LIC-adapted pET-28a

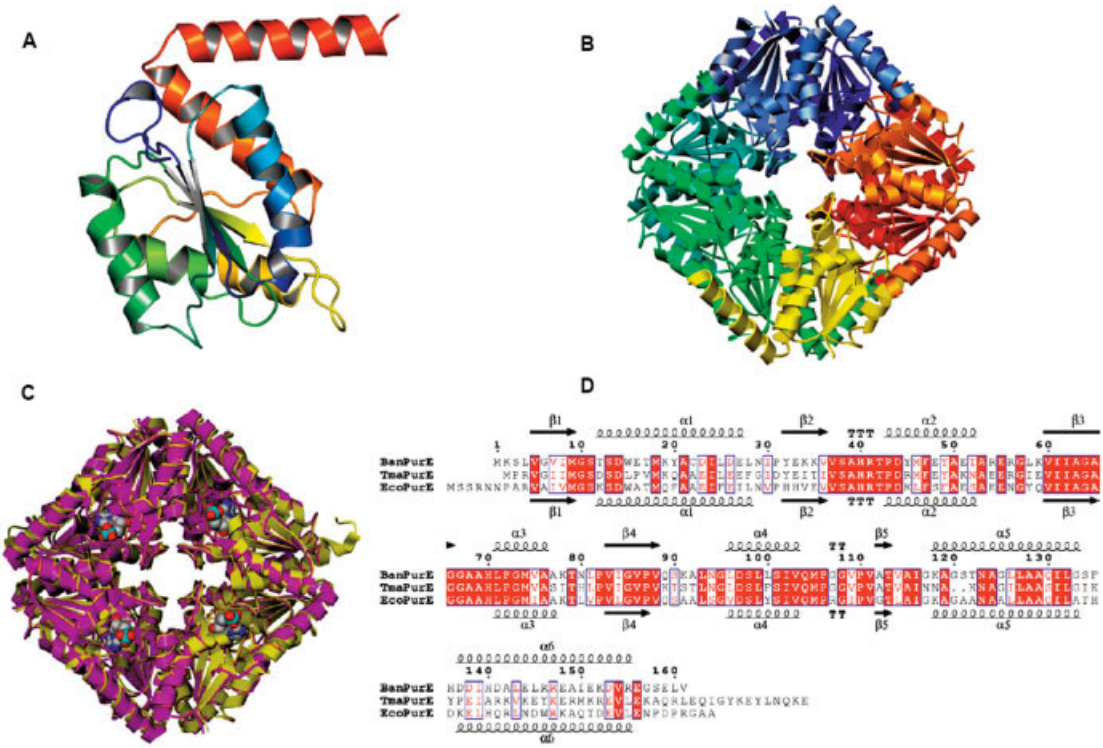

Fig. 1. Features of the PurE structure from *Bacillus anthracis*: (**A**) ribbon tracing of a PurE subunit color-ramped from the amino-terminus (blue) to the carboxyl-terminus (red); (**B**) the octamer colored by subunit; (**C**) superposition of Ban (yellow) and Eco (magenta) PurE structures, with (N^6-CAIR) atoms[6] represented as space filling; (**D**) sequence alignment generated using ESPript[17] with Ban (top) and Eco (bottom) PurE secondary-structure elements superimposed. All structure images were made by the built-in ray-tracing libraries of pyMOL (De Lano Scientific) or using MOLSCRIPT[18] and RASTER3D[19] for rendering.

(pET-YSBLIC; MJF, manuscript in preparation) by cleavage with the restriction endonuclease *Bse*RI and treatment with T4 DNA polymerase in the presence of 2′-deoxythymidine 5′-triphosphate (dTTP). The vector and PCR products were annealed and used to transform *E. coli* NovaBlue cells (Novagen) to kanamycin resistance. Colony PCR using T7-promoter and gene-specific oligonucleotide primers confirmed the presence of gene insert, whose sequence was then confirmed. The pET-YSBLIC*purE* clone was transformed into *E. coli* BL21 (DE3) for protein overexpression.

A 1 L culture of cells was grown with shaking at 37°C in Luria-Bertani broth containing 30 μg/mL kanamycin to an OD_{600} of 0.6, at which point expression of the *B. anthracis* full-length protein PurE, fused to an amino terminal 6-His tag (MGSSHHHHHH) encoded by pET-YSBLIC, was induced by the addition of isopropyl-β-D-thiogalactopyranoside to a final concentration of 1 m*M*. Cells were harvested by centrifugation, resuspended in 20 m*M* Na_2HPO_4 buffer (pH 7.5) containing 500 m*M* NaCl and 10 m*M* imidazole (Buffer A), lysed by sonication on ice, and the cell-free supernatant collected by centrifugation at 20,000 rpm in a Sorvall SS-34 rotor (DuPont). A 5 mL Hi-Trap chelating column (Amersham Pharmacia), previously charged with nickel and equilibrated with Buffer A, was attached to an ÄKTA Explorer 3D purification system (Amersham Pharmacia). The soluble fraction was applied to the column and washed with 25 mL Buffer A followed by 50 mL of Buffer A containing 70 m*M* imidazole. The protein was eluted with Buffer A containing 500 m*M* imidazole and automatically passed to a Superdex G-200 gel filtration column (Amersham Pharmacia) pre-equilibrated in 50 m*M* Tris buffer (pH 7.5) containing 150 m*M* NaCl for size exclusion chromatography. Fractions containing PurE were identified by SDS-PAGE, pooled, and concentrated to 21 mg/mL by centrifugal ultrafiltration (Millipore).

Automated crystallization screening was carried out in 96-well plates (Greiner) using a Mosquito nanoliter pipetting robot (TTP labtech) and screens from Hampton Research (I, II, and Index), and Clear Strategy Screens (CSSI and CSSII)[14] at pH 7.5 and 6.5, using sitting-drop vapor diffusion. Each crystallization drop contained 150 nL of protein solution and 150 nL of reservoir solution. The protein crystallized from a reservoir solution containing 15% polyethylene glycol (PEG) 4000, 0.8*M* sodium formate and 0.1*M* Tris-HCl pH 7.5.

Data collection. Native diffraction data were collected on beamline PX 9.6 at the Daresbury Synchrotron Radiation Source (SRS). A single crystal was mounted in a rayon loop and transferred to a solution of the crystallization mother liquor supplemented with 30% PEG 4000 and then rapidly cooled in liquid nitrogen. Diffraction data were recorded on an ADSC Quantum 4 CCD detector. Data were processed and reduced using the HKL2000 software suite (Denzo and Scalepack). Data collection and refinement statistics are presented in Table I.

Structure Solution and Refinement. The structure of PurE was determined by molecular replacement using the program MOLREP[15] from the CCP4 suite.[16] Data in the resolution range 50–4 Å were used in both rotation and translation calculations. Refinement calculations were performed using REFMAC5,[8] interspersed with sessions of manual modelling using QUANTA. Refinement and model statistics are presented in Table I.

Data validation and deposition. The quality of the final model was scrutinized using PROCHECK and SFCHECK.[7] Atomic coordinates for the refined *B. anthracis* PurE model, together with experimental structure factors, have been deposited as PDB code 1XMP.

ACKNOWLEDGMENTS

We thank Professor Colin Harwood (Newcastle) for providing genomic DNA, and the SRS Daresbury for synchrotron beam time. The work described here was funded by the European Commission as SPINE, contract-no. QLG2-CT-2002-00988 under the RTD program "Quality of Life and Management of Living Resources." MPB and VML are supported by the BBSRC, JAB by the Wellcome Trust.

REFERENCES

1. Tiedeman AA, Keyhani J, Kamholz J, Daum HA, Gots JA, Smith JM. Nucleotide sequence analysis of the the *purEK* operon encoding 5′-phosphoribosyl-5-aminoimidazole carboxylase of *Escherichia coli* K-12. J Bacteriol 1989;171:205–212.
2. Mueller EJ, Meyer E, Rudolph J, Davisson VJ, Stubbe J. N^5-carboxyaminoimidazole ribonucleotide: evidence for a new intermediate and two new enzymatic activities in the *de novo* purine biosynthetic pathway of *Escherichia coli*. Biochemistry 1994;33:2269–2278.
3. Meyer E, Kappock TJ, Osuji C, Stubbe J. Evidence for the direct transfer of the carboxylate of N^5-carboxyaminoimidazole ribonucleotide (N^5-CAIR) to generate 4-carboxy-5-aminoimidazole ribonucleotide catalyzed by *Escherichia coli* PurE, an N^5-CAIR mutase. Biochemistry 1999;38:3012–3018.
4. Firestine SM, Poon S-W, Mueller EJ, Stubbe J, Davisson VJ. Reactions catalysed by 5-aminoimidazole ribonucleotide carboxylases from *Escherichia coli* and *Gallus gallus*: a case for divergent catalytic mechanisms? Biochemistry 1994;33:11927–11934.
5. Firestine SM, Davisson VJ. A tight binding inhibitor of 5-aminoimidazole ribonucleotide carboxylase. J Med Chem 1993;36:3484–3486.
6. Matthews II, Kappock TJ, Stubbe J, Ealick SE. Crystal structure of *Escherichia coli* PurE, an unusual mutase in the purine biosynthetic pathway. Structure 1999;7:1395–1406.
7. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
8. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53:240–255.
9. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
10. Schwarzenbacher R, Jaroszewski L, von Delft F, Abdubek P, Biorac T, Brinen LS, et al. Crystal structure of a phosphoribosylaminoimidazole mutase PurE (TM0446) from *Thermotoga maritima* at 1.77-Å resolution. Proteins 2004;55:474–478.
11. Holm L, Sander C. DALI: A network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
12. Lukens LN and Buchanan JM. Biosynthesis of the purines. xxvi. The enzymatic synthesis of 5-amino-1-ribosyl-4-imidazolecaboxylic acid 5′-phosphate from 5-amino-1-ribosylimidazole 5′-phosphate and carbon dioxide. J Biol Chem 1959;234:1799–1805.
13. Aslandis C, De Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). Nuc Acid Res 1990;18:6069–6074.
14. Brzozowski AM, Walton J. Clear strategy screens for macromolecular crystallisation. J Appl Crystallogr 2001;34:97–101.
15. Vagin A, Teplyakov A. MOLREP: An automated program for molecular replacement. J Appl Crystallogr 1997;30:1022–1025.
16. Collaborative Computational Project No.4. The CCP4 suite: Programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
17. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: multiple sequence alignments in PostScript. Bioinformatics 1999;15:305–308.
18. Kraulis PJ. Molscript: A program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:946–950.
19. Merritt EA, Bacon DJ. RASTER3D: Photorealistic molecular graphics. Meth Enzymol 1997;277:505–524.