Structure of purine nucleoside phosphorylase (DeoD) from *Bacillus anthracis*

Rosa Grenha, Vladimir M. Levdikov, Mark J. Fogg, Elena V. Blagova, James A. Brannigan,* Anthony J. Wilkinson and Keith S. Wilson

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

Correspondence e-mail: jab@ysbl.york.ac.uk

Protein structures from the causative agent of anthrax (*Bacillus anthracis*) are being determined as part of a structural genomics programme. Amongst initial candidates for crystallographic analysis are enzymes involved in nucleotide biosynthesis, since these are recognized as potential targets in antibacterial therapy. Purine nucleoside phosphorylase is a key enzyme in the purine-salvage pathway. The crystal structure of purine nucleoside phosphorylase (DeoD) from *B. anthracis* has been solved by molecular replacement at 2.24 Å resolution and refined to an *R* factor of 18.4%. This is the first report of a DeoD structure from a Gram-positive bacterium.

1. Introduction

Purine nucleoside phosphorylase (PNP) is involved in the purine-salvage pathway. It promotes the reversible phosphorolysis of the glycosidic bond of purine ribo- and deoxyribonucleosides and their analogues. This phosphorolysis by inorganic phosphate produces the purine base and (deoxy)ribose-1-phosphate as products. Previous studies have shown there are two distinct classes of nucleoside phosphorylases (Pugmire & Ealick, 2002). Members of the NP-I (nucleoside phosphorylase I) family are either trimers or hexamers made up of subunits that contain an α/β-fold. Proteins of the NP-II family are dimers in which each subunit has a small α-domain separated from an α/β-domain by a large crevice.

Enzymes belonging to the NP-I family can be further classified according to their substrate specificity and amino-acid sequences. Trimeric PNPs specific for guanine and hypoxanthine (2′-deoxy)-ribonucleosides are present in mammalian species. The hexameric PNPs, which accept a broader range of substrates including adenosine, are prevalent in bacterial species, although *Escherichia coli* (Seeger *et al.*, 1995), *Bacillus subtilis* (Senesi *et al.*, 1976) and *B. stearothermophilus* (Hori *et al.*, 1989) appear to have both trimeric and hexameric forms.

B. anthracis is a Gram-positive rod-shaped spore-forming bacterium. It is the causative agent of anthrax, which is an acute, infectious and normally lethal disease. The *B. anthracis* genome was sequenced at The Institute for Genomic Research (TIGR; Read *et al.*, 2003). As in other *Bacillus* species, there are two gene products annotated as PNPs in the genomic sequence database for *B. anthracis* strain Ames. We have solved the structure of one of the *B. anthracis* enzymes, hereafter referred to as DeoD, which is designated BA1483 in the TIGR database (http://www.tigr.org).

2. Materials and methods

2.1. Cloning, expression and purification

The *deoD* gene encoding BA1483 was amplified by the polymerase chain reaction (PCR) from *B. anthracis* genomic DNA using the primers 5′-**CACCACCACCAC**ATGAGTGTACATATTGAAGCA-AAACAAGGCG-3′ and 5′-**GAGGAGAAGGCGCGTTA**TTGTTG-AATTGCTGCATCTAAAGCGATTTC-3′ to incorporate sequences (in bold) designed to generate compatible overhangs with the vector pETYSBLIC. This plasmid is a pET28a derivative that has been
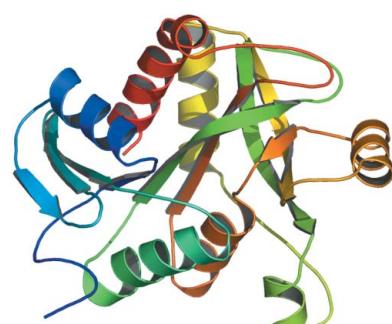
Table 1
X-ray data-collection and refinement statistics.

Values in parentheses represent data for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å, °)	$a = 63.86$, $b = 128.26$, $c = 223.57$, $\alpha = \beta = \gamma = 90$
Resolution range (Å)	20.00–2.24 (2.32–2.24)
No. of unique reflections	81257 (7218)
Completeness (%)	90.8 (82.0)
Redundancy	3.9 (3.2)
$I/\sigma(I)$	8.14 (1.84)
R_{merge} (%)	14.1 (49.1)
R factor†	0.184 (0.239)
R_{free}† (5% of diffraction data)	0.235 (0.308)
No. of subunits in the asymmetric unit	6
No. of protein non-H atoms	11450
No. of water molecules	716
R.m.s.d. bond lengths (Å)	0.014
R.m.s.d. angles (°)	1.538
Mean B value (Å^2)	
Solvent	25.5
Main chain	21.7
Side chain	23.4

† R factor $= \sum |F_{obs}| - |F_{calc}|/\sum |F_{obs}|$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. R_{free} is the R factor calculated with 5% of the reflections chosen at random and omitted from refinement.

adapted for ligation-independent cloning (LIC; Aslanidis & De Jong, 1990) and confers an N-terminal hexa-histidine tag to the cloned gene. The reaction products were used to transform *E. coli* NovaBlue competent cells (Novagen) and selected through kanamycin resistance. Single-colony PCR screening and DNA sequencing allowed the isolation of the desired recombinants. The recombinant plasmid pETYSBL*deoD* was introduced into *E. coli* strain B834 (DE3) and grown at 310 K in 1 l Luria–Bertani broth supplemented with 35 µg ml^{-1} kanamycin to an OD$_{600}$ of 0.5. Expression of DeoD was then induced by addition of isopropyl-β-D-thiogalactopyranoside to 1 mM and the cells were grown for a further 3 h at 310 K. Cells were harvested by centrifugation at 6080g for 10 min (Sorvall GS-3) and frozen overnight at 193 K. The pellet was thawed and resuspended in buffer A (20 mM Na$_2$HPO$_4$, 0.5 M NaCl, 10 mM imidazole pH 7.5) and lysed by sonication (Soniprep 150, MSE). Cell debris was removed by centrifugation at 26 890g for 60 min (Sorvall SS34) and the supernatant cleared by filtration through a 0.2 µm device (Sartorius), yielding 10 ml crude extract. The supernatant was loaded onto a 5 ml HiTrap nickel-chelating column (Äkta Explorer system, Amersham Biosciences) previously equilibrated with buffer A. The protein was eluted in buffer A containing 500 mM imidazole and automatically directed onto a HiLoad 16/60 Superdex 200 prep-grade gel-filtration column (Amersham Biosciences) pre-equilibrated in 50 mM Tris–HCl pH 7.5 and 250 mM NaCl. The purified protein was concentrated to 6.2 mg ml^{-1} in 25 mM NaCl, 25 mM Tris–HCl pH 7.5.

2.2. Protein characterization

Electrospray mass spectrometry performed on an ABI Qstar tandem mass spectrometer using 20 µl of a 4.5 mg ml^{-1} solution of purified protein dialysed against 2 mM Tris–HCl pH 7.5 buffer showed a peak (26 733 Da) in close agreement with the calculated DeoD molecular weight. The circular-dichroism spectrum was recorded between 180 and 260 nm on a Jasco J810 CD Spectrophotometer using 400 µl pure protein solution (0.2 mg ml^{-1}) in 25 mM Tris–HCl pH 7.5 buffer containing 25 mM NaCl. Spectrum analysis using the program *K2d* (Andrade *et al.*, 1993) suggested a predominance (51%) of α-helical content, plus 16% β-sheet and 33% random coil.

2.3. Crystallization

Preliminary crystallization screening was performed by sitting-drop vapour diffusion with nanodrops (150 nl, equilibrated against 80 µl reservoir solution) using a Mosquito Nanolitre Pipetting robot (TTP LabTech) to set up 96-well plates at 298 K. Four commercial screens were used: Crystal Screens I, II and Index (Hampton Research), and Stura Footprint (Molecular Dimensions). Crystals grew overnight from a number of different conditions. These included orthorhombic shaped crystals from 2.0 M ammonium sulfate plus 5% 2-propanol or 30%(v/v) PEG 400, 0.1 M HEPES pH 7.5, 0.2 M
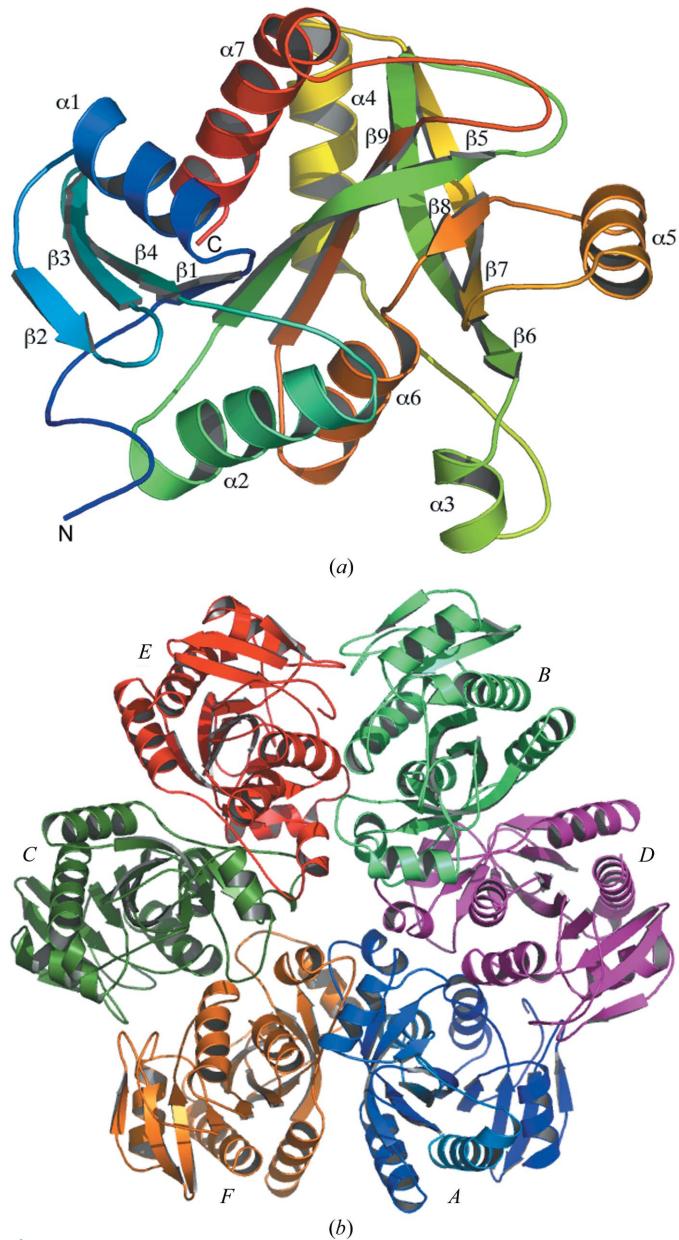
Figure 1
(*a*) Ribbon diagram of a single subunit of *B. anthracis* DeoD. The α-helices and β-strands are labelled in sequential order. The chain is ramped in rainbow colours from the N-terminus (blue) to the C-terminus (red). The structure figures were drawn with *PyMol* (DeLano, 2002). (*b*) Ribbon diagram of purine nucleoside phosphorylase. There are two types of interactions between subunits. Between *A* and *F*, *B* and *D*, and *C* and *E* there are extensive interactions to form dimers through amino acids on helices α3 and α5, strands β6 and β7 and loops α3–α4, β8–α6 and α3–β6. Interactions between subunits *A* and *D*, *C* and *F*, and *B* and *E* are based on the contribution of amino acids from helices α1, α2 and α3 and loops β2–β3, α2–β4 and α5–β7; these build the dimers into a hexamer.

MgCl$_2$·6H$_2$O solutions. In addition, hexagonal plates from 2.0 *M* ammonium sulfate, 2%(*v*/*v*) PEG 400, 0.1 *M* HEPES pH 7.5 or 2.4 *M* sodium malonate pH 7.0 solutions were obtained. Crystal growth was scaled up by hanging-drop vapour diffusion at 293 K in 24-well plates using a 1:1 mixture of protein solution (6.2 mg ml^{-1}) and reservoir solution (1 µl each). For diffraction data collection, an orthorhombic crystal of length 80 µm was harvested after overnight growth from 25%(*v*/*v*) PEG 400, 0.1 *M* HEPES pH 7.5, 0.2 *M* MgCl$_2$·6H$_2$O.

2.4. Data collection and processing

A single crystal was mounted in a loop and flash-cooled in liquid nitrogen using the mother liquor as cryoprotectant. X-ray diffraction data were collected in-house at 120 K on an image-plate detector equipped with a MAR Research 345 scanner to a resolution of 2.24 Å using a Rigaku RUH3R X-ray generator. The crystal-to-detector distance was 150 mm and images consisting of 0.5° rotations were exposed for 10 min. X-ray diffraction data were processed using *DENZO* and *SCALEPACK* from the *HKL*2000 package within the *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994). Data statistics are shown in Table 1.

## 2.5. Structure solution and refinement

The structure was solved by molecular replacement with *MOLREP* (Vagin & Teplyakov, 1997) using the homologous (56% sequence identity) *E. coli* PNP structure (PDB code 1ecp) as a starting model. Data in the resolution range 20–3 Å were used in both rotation and translation calculations, which gave an obvious solution with significant contrast, resulting in six molecules in the asymmetric unit and a solvent content of 58.4%. After a rigid-body refinement performed by *REFMAC* (Murshudov *et al.*, 1997), the correlation coefficient was 59.6% and the *R* factor was 45.7% (R_{free} = 45.6%). Structure refinement was continued using *REFMAC* and model building in *QUANTA* (Accelrys Inc., San Diego, CA, USA) and *COOT* (Emsley & Cowtan, 2004). The final *R* factor for the model consisting of 1399 amino-acid residues and 716 water molecules is 18.4% (R_{free} = 23.5%). Refinement statistics are listed in Table 1. The structure was validated with *PROCHECK*3.5.4 (Laskowski *et al.*, 1993) and *SFCHECK*6.0.3 (Vaguine *et al.*, 1999).

3. Results and discussion

The refined model of *B. anthracis* DeoD is composed of six virtually identical subunits (*A*–*F*) in the asymmetric unit. Most residues are in well defined electron density, except for the His$_6$ tag. Residues 1–232 (numbering from the native protein initiation methionine) are visible in chains *A*, *B* and *C*, 1–233 in chains *D* and *F* and 1–234 in chain *E*, out of a total of 235 in the polypeptide chain. Analysis of the Ramachandran plot showed that 91.7% of the non-glycine residues are in the favoured regions, 8.1% in the additional allowed regions and 0.2% in the generously allowed regions.

Each subunit has an α/β-fold consisting of nine β-strands in a highly twisted mixed β-sheet flanked by seven α-helices carrying several extended loops (Fig. 1*a*). The quaternary structure is similar to that of bacterial PNPs, *e.g. E. coli* (Mao *et al.*, 1997) and *Thermus thermophilus* (Tahirov *et al.*, 2004). It consists of a disc-shaped hexamer with *D*3 symmetry, organized as a trimer of dimers with approximate dimensions 101 × 96 × 42 Å. The interactions between subunits forming each dimer are substantially more extensive than those between dimers (Fig. 1*b*).

Sequence alignments with other hexameric PNPs show a high sequence identity (∼60%) and superimposition with other already deposited structures reveals a highly conserved subunit topology and active-site location (Fig. 2). Previous studies have shown that hexameric PNPs have two active sites per dimer. The DeoD structure has wide and accessible major grooves between the two subunits forming each dimer. Superimposition of DeoD with PNP structures containing a ligand shows that these grooves usually accommodate the substrate. There are several conserved residues in this region (Fig. 3*a*). Residue His4 contributes to ribose binding and Arg43 balances the negative charge on the phosphate. These two residues are recruited from one subunit and, together with a number of residues from the adjacent subunit, are implicated in ligand binding to the active site (Fig. 3*b*). These include Arg87 and Thr90, which also interact with the phosphate (Mao *et al.*, 1997). Binding of the base moiety is predominantly *via* a stacking interaction with Phe159 and base specificity is conferred by Asp204 (Tahirov *et al.*, 2004). Glu181 forms a bidentate interaction with the ribose O2′- and O3′-hydroxyl groups. This organization of subunit cooperation leads to three active sites on each face of the hexamer.
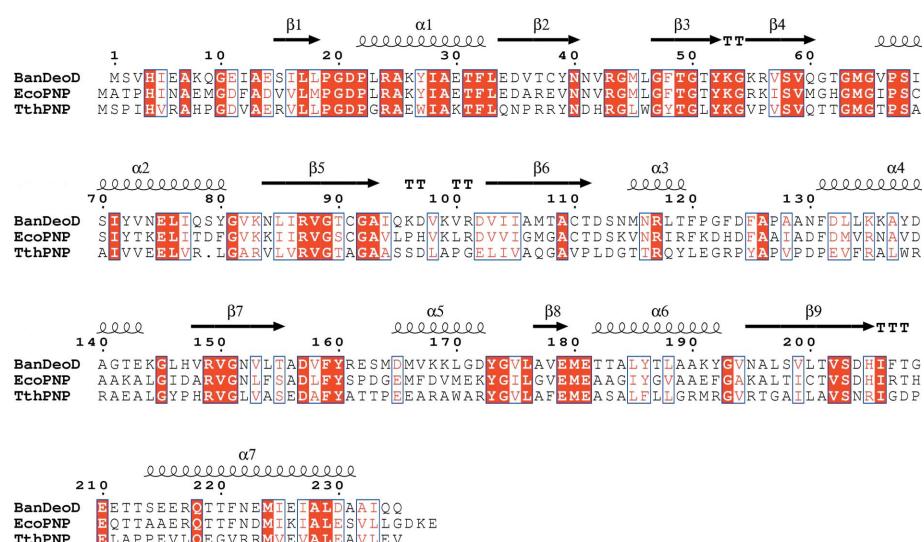
Figure 2
Sequence alignment of *B. anthracis* DeoD with purine nucleoside phosphorylases from *E. coli* (Eco) and *T. thermophilus* (Tth). The sequence-alignment figure was generated using *ESPript* (Gouet *et al.*, 1999). Strictly conserved residues are highlighted with red boxes; conservative substitutions are also boxed.
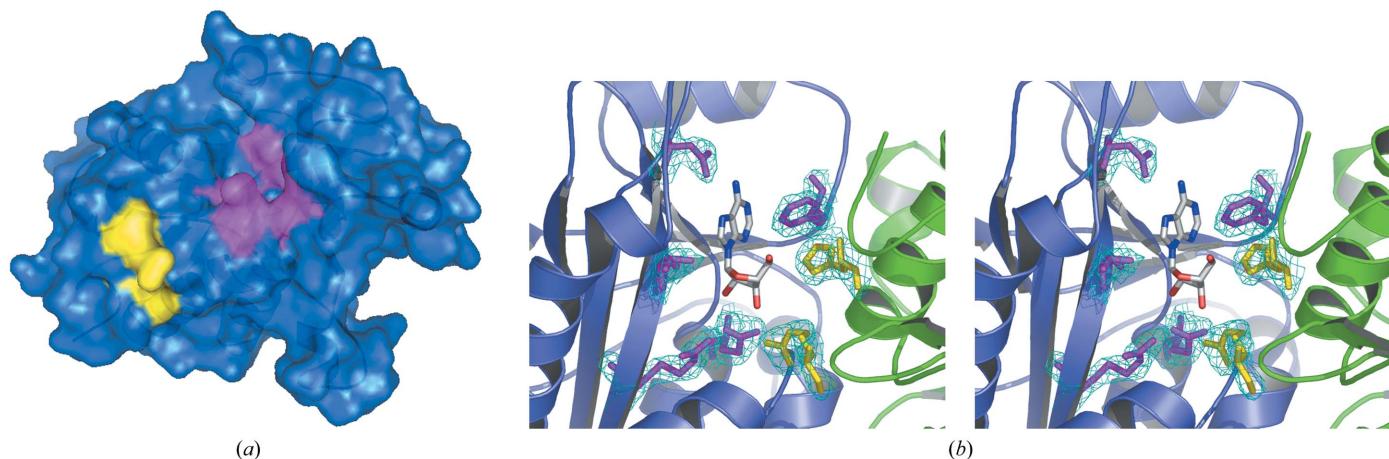
Figure 3
(*a*) Surface representation of a single subunit, showing the pocket that accommodates the substrate (magenta) that is augmented by interactions with residues His4 and Arg43 (yellow) from the neighbouring subunit. (*b*). Stereoview of ligand binding, based on the position of adenosine (stick representation, coloured by atom) in *T. thermophilus* PNP (Tahirov *et al.*, 2004). The electron density for selected side chains is represented as wire contoured at a level of 1σ. These are (clockwise from bottom of figure) Glu181, Arg87, Thr90, Asp204 and Phe159 in magenta from one subunit (blue). His4 and Arg43 (yellow) from the adjacent subunit (green) complete the active centre (see text).

There are differences in substrate specificity between human and bacterial PNPs. These have been exploited in potential anticancer therapies, where a gene for a bacterial PNP is transfected into a tumour and the protein converts a non-toxic pro-drug to a cytotoxic species (Sorscher *et al.*, 1994). Further analysis of the structure of *B. anthracis* DeoD and its complexes may reveal differences that would allow selective inhibition of DeoD activity.

References

Andrade, M. A., Chacon, P., Merelo, J. J. & Moran, F. (1993). *Protein Eng.* **6**, 383–390.
Aslanidis, C. & De Jong, P. J. (1990). *Nucleic Acids Res.* **18**, 6069–6074.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
DeLano, W. L. (2002). *PyMol.* DeLano Scientific, San Carlos, CA, USA.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Hori, N., Watanabe, M., Yamazaki, Y. & Mikami, Y. (1989). *Agric. Biol. Chem.* **53**, 2205–2210.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Mao, C., Cook, W. J., Zhou, M., Koszalka, G. W., Krenitsky, T. A. & Ealick, S. E. (1997). *Structure*, **5**, 1373–1383.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Pugmire, M. J. & Ealick, S. E. (2002). *Biochem. J.* **361**, 1–25.
Read, T. D. *et al.* (2003). *Nature (London)*, **423**, 81–86.
Seeger, C., Poulsen, C. & Dandanell, G. (1995). *J. Bacteriol.* **177**, 5506–5516.
Senesi, S., Falcone, G., Mura, U., Sgarrella, F. & Ipata, P. L. (1976). *FEBS Lett.* **64**, 353–357.
Sorscher, E. J., Peng, S., Bebok, Z., Allan, P. W., Bennett, L. L. Jr & Parker, W. B. (1994). *Gene Ther.* **1**, 233–238.
Tahirov, T. H., Inagaki, E., Ohshima, N., Kitao, T., Kuroishi, C., Ukita, Y., Takio, K., Kobayashi, M., Kuramitsu, S., Yokoyama, S. & Miyano, M. (2004). *J. Mol. Biol.* **337**, 1149–1160.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.