

Satyabrata Das, Parimal Kumar,  
Vikrant Bhor, A. Surolia\* and  
M. Vijayan\*Molecular Biophysics Unit, Indian Institute of  
Science, Bangalore 560 012, IndiaCorrespondence e-mail:  
surolia@mbu.iisc.ernet.in,  
mv@mbu.iisc.ernet.in

Received 22 August 2004

Accepted 1 November 2004

Online 9 November 2004

## Expression, purification, crystallization and preliminary X-ray crystallographic analysis of pantothenate kinase from *Mycobacterium tuberculosis*

Pantothenate kinase is an essential enzyme in the bacterial life cycle. It catalyzes the phosphorylation of pantothenate (vitamin B<sub>5</sub>) to 4'-phosphopantothenate, the first step in the coenzyme A biosynthetic pathway. The enzyme from *Mycobacterium tuberculosis*, MW 35.7 kDa, has been cloned, expressed, purified and crystallized in two different trigonal crystal forms, both belonging to space group *P*3<sub>1</sub>21. Two complete data sets of resolution 2.5 Å (form I) and 2.9 Å (form II) from crystals with unit-cell parameters  $a = b = 78.3$ ,  $c = 115.45$  Å and  $a = b = 107.63$ ,  $c = 89.85$  Å, respectively, were collected at room temperature on a home X-ray source. Structures of both crystal forms were solved for one subunit in the asymmetric unit by molecular replacement.

### 1. Introduction

Pantothenate kinase (PanK) is a ubiquitous and essential enzyme that catalyzes the first step of the universal coenzyme A (CoA) biosynthetic pathway. In this step, pantothenate (vitamin B<sub>5</sub>) is converted to 4'-phosphopantothenate, which subsequently forms CoA by four enzymatic steps (Begley *et al.*, 2001). CoA is an important cofactor and it interacts with almost 4% of known enzymes with very diverse folds and function (Engel & Wierenga, 1996). The PanK-catalyzed reaction is the rate-controlling step in CoA biosynthesis and is subject to regulation by feedback inhibition by CoA and its thioesters (Rock *et al.*, 2003). Although PanK occurs universally in living organisms, the sequence similarity between the prokaryotic and the eukaryotic enzymes is very low (Rock *et al.*, 2000). It has been shown that defects in the *PanK2* gene in humans are associated with neurodegenerative disorders, formerly known as Hallervorden–Spatz syndrome (Zhou *et al.*, 2001).

Presently, the only known structure of PanK is that from *Escherichia coli*, EcPanK (Yun *et al.*, 2000), in complex with CoA or with 5'-adenylymidodiphosphate (AMPPNP), a non-hydrolysable analogue of ATP. EcPanK is a homodimeric protein. Each subunit (MW 36 kDa) consists of 316 amino acids. It has a three-layer  $\alpha\beta\alpha$  architecture in which a seven-stranded  $\beta$ -sheet is flanked by two helices on each side. An extensive antiparallel coiled coil formed by two long helices is present in the dimerization interface. EcPanK has a phosphate-binding loop (known as the P-loop) with the sequence signature GXXXXGKS as observed in other kinases (Leipe *et al.*, 2003). Kinetic studies show that the binding of ATP to EcPanK is highly cooperative in nature and that CoA inhibits the activity by competitive binding to the ATP site (Calder *et al.*, 1999). This cooperativity is presumably related to the dimeric nature of the enzyme. The interaction of AMPPNP and CoA with EcPanK observed in the crystal structures takes place in very different ways; however, their phosphate group interacts with Lys101 in the P-loop, which explains the kinetic competition between the regulator CoA and the substrate ATP.

The Gram-positive bacterium *Mycobacterium tuberculosis* (Mt) causes the infectious disease tuberculosis in humans. The complete genomic sequence of the best-characterized laboratory strain of *M. tuberculosis*, H37Rv, is available (Cole *et al.*, 1998). A worldwide effort is presently taking place to determine the structures of Mt proteins (Goulding *et al.*, 2002; Terwilliger *et al.*, 2003). The structures of several proteins from this pathogen and the homologous



*M. smegmatis* have already been reported from this laboratory (Datta *et al.*, 2000, 2003; Saikrishnan *et al.*, 2003; Roy *et al.*, 2004). Here, we report the cloning, expression, purification, crystallization and structure solution of PanK from *M. tuberculosis* (MtPanK). The dimeric protein has a 312-residue subunit with MW 35.7 kDa. MtPanK is homologous to EcPanK and shows 51% sequence identity.

## 2. Materials and methods

### 2.1. Cloning, overexpression and purification

The PanK gene was amplified using the genomic DNA of *M. tuberculosis* H37Rv and the oligonucleotide primers 5'-CATGC-CATGGCATCATCATCATCATATGTCTCGCGGCTTAGCGA-GCCGAGCCCA-3' and 5'-CCGGAATTCTTACAGCTTGCGAG-CCGAGCCGGT-3'. The amplified gene was inserted into the *Nco*II/*Eco*RI-digested expression vector pET-28a(+) (Novagen) such that six histidine residues were added to the N-terminus of the gene product in order to facilitate protein purification. The recombinant plasmid was transformed into *Escherichia coli* BL21 (DE3) cells (Novagen) and cultures were grown in Luria-Bertani medium supplemented with 30 µg ml<sup>-1</sup> kanamycin at 310 K until *A*<sub>600</sub> reached 0.6; cultures were then induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated for a further 7 h. The cells were harvested by centrifugation at 9190g (Sorvall SLC-4000 rotor) for 10 min at 277 K. The cell pellet was resuspended in ice-cold lysis buffer [0.025 M Tris-HCl pH 8.0, 0.6 M NaCl, 0.025 M imidazole, 5% (v/v) glycerol] and lysed by sonication. The crude lysate was centrifuged at 26 892g (Sorvall SS34 rotor) for 30 min at 277 K and the recombinant protein was purified from the supernatant fraction in a single step using a Ni-NTA metal-affinity column (His-bind resin, Novagen). The purity of the protein was judged by SDS-PAGE analysis.

### 2.2. Crystallization

The initial crystallization screening was performed using commercially available sparse-matrix screens (Jancarik & Kim, 1991) from Hampton Research, employing the hanging-drop vapour-diffusion method. All crystallization trials were performed at room temperature (293 K) using 24-well Linbro tissue-culture plates. The conditions were varied in terms of the concentration of the protein



**Figure 1**  
Crystal of pantothenate kinase from *M. tuberculosis*; crystals of both forms are similar in shape.

**Table 1**

Crystal data and data-collection statistics.

Values in parentheses are for the highest resolution shell.

	Form I	Form II
Space group	<i>P</i> 3 <sub>1</sub> 21	<i>P</i> 3 <sub>1</sub> 21
Unit-cell parameters (Å)		
<i>a</i>	78.3	107.63
<i>c</i>	115.45	89.85
Resolution range (Å)	20.0–2.5 (2.59–2.5)	20.0–2.9 (3.0–2.9)
Matthews coefficient <i>V</i> <sub>M</sub> (Å <sup>3</sup> Da <sup>-1</sup> )	2.8	4.1
Solvent content (%)	55.6	69.8
Subunits per AU	1	1
Total No. reflections measured	71825	62923
No. unique reflections	14530 (1442)	13523 (1340)
Average multiplicity	4.9 (5.0)	4.7 (4.7)
Data completeness (%)	99.2 (100.0)	99.0 (99.7)
Mean <i>I</i> / <i>σ</i> ( <i>I</i> )	15.1 (3.9)	14.0 (3.3)
<i>R</i> <sub>merge</sub> <sup>†</sup> (%)	7.3 (49.0)	8.8 (53.6)

<sup>†</sup>  $R_{\text{merge}} = \sum_h \sum_i |I(h, i) - \langle I(h) \rangle| / \sum_h \sum_i I(h, i)$ , where *I*(*h*, *i*) is the intensity of the *i*th observation of reflection *h* and  $\langle I(h) \rangle$  is the mean value of *I*(*h*, *i*) for all measurements.

and glycerol and the Hampton solutions until diffraction-quality crystals appeared.

### 2.3. X-ray data collection and structure analysis

X-ray diffraction data sets were recorded at 293 K using an in-house MAR345 imaging-plate system mounted on a Rigaku Ultrax-18 rotating-anode X-ray generator operating at 40 kV and 80 mA with a copper anode. The X-ray beams were focused using Osmic mirrors. All images were recorded with an oscillation angle of 1°, an exposure time of 300 s per frame and a crystal-to-detector distance of 120 mm. The data were processed and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). The intensity data thus obtained were truncated to amplitudes using the program *TRUNCATE* (French & Wilson, 1978) from the *CCP4* program package (Collaborative Computational Project, Number 4, 1994) and used for structure solution and refinement. In order to estimate the protein content of the asymmetric unit, the Matthews coefficient (*V*<sub>M</sub>; Matthews, 1968) and solvent content were calculated based on a subunit MW of 36.6 kDa (including six histidine residues added to the N-terminus of MtPanK during cloning). *AMoRe* (Navaza, 1994) was used for all molecular-replacement calculations and *CNS* v.1.1 (Brünger *et al.*, 1998) was used for refinement.

## 3. Results and discussion

The MtPanK gene with an N-terminal His<sub>6</sub> tag was cloned and expressed in *E. coli* BL21 (DE3) cells. A Ni-NTA affinity column was used for purification and the homogeneity of the purified sample was verified by SDS-PAGE analysis. The enzyme was crystallized in two trigonal crystal forms of similar shape: forms I and II (Fig. 1). Crystallization droplets were prepared by mixing 3 µl protein solution (6 mg ml<sup>-1</sup>) in 0.1 M Tris-HCl, 0.15 M NaCl, 5% (v/v) glycerol and 0.001 M β-mercaptoethanol at pH 8.0 with 1 µl precipitant solution consisting of 10–15% (w/v) PEG 8000, 0.05–0.1 M NaOAc and 0.05 M NaCl dissolved in 0.1 M sodium cacodylate buffer pH 6.5 and were equilibrated against 400 µl of the same precipitant solution. Diffraction-quality crystals of both forms appeared in 3 d and grew to approximate dimensions of 0.5 × 0.4 × 0.2 mm (Fig. 1) within a week. The crystal data and data-collection statistics are listed in Table 1. Systematic absences indicated the space group in both cases to be *P*3<sub>1</sub>21 or *P*3<sub>2</sub>21. The latter was subsequently confirmed by structure solution.

Molecular-replacement (MR) calculations using *AMoRe* did not lead to structure solutions in space group  $P3_221$ , but did in space group  $P3_121$ . A 311-residue polyaniline chain of the EcPanK structure (PDB code 1esm) was used as the search model. The correlation coefficient (CC) and  $R$  factor for the correct unrefined solution were 0.43 and 0.47, respectively, for 10.0–3.0 Å data in the case of form I. CC and  $R$  for the second best solution were 0.34 and 0.50, respectively. The corresponding values in the case of crystal form II were 0.53 and 0.46 (0.42 and 0.50 for the second best solution). In both cases, initial  $\sigma$ -weighted electron-density maps with Fourier coefficients  $2F_o - F_c$  and MR phases were of interpretable quality almost throughout the chain. Iterative cycles of model building using *FRODO* (Jones, 1978) and refinement using *CNS* v.1.1 (Brünger *et al.*, 1998) lowered the  $R$  and  $R_{\text{free}}$  values to 0.28 and 0.32 for form I and 0.29 and 0.32 for form II, respectively. Further refinement and model building are in progress.

The structure of the molecule in the two crystal forms is nearly the same except for a difference of about 6° in the mutual orientation of the two subunits. The structure is also grossly similar to that of EcPanK. However, even at the present stage of refinement, several differences between the two enzymes are discernable. In particular, major differences exist in the intersubunit interface and in the mutual orientation of the two subunits. The implications of these differences, particularly for cooperativity, merit further exploration.

The intensity data were collected using the X-ray Facility for Structural Biology at the Institute, supported by the Department of Science and Technology (DST) and the Department of Biotechnology (DBT), Government of India. Computations and model building were performed at the Supercomputer Education and Research Centre and the DBT-supported Bioinformatics Centre and Graphics Facility. Support from the Fund for Applied Research of the Centre for Scientific and Industrial Consultancy to this Institute is

acknowledged. MV is supported by a Distinguished Biotechnologist Award of the DBT. SD is a DBT postdoctoral fellow.

## References

- Begley, T. P., Kinsland, C. & Strauss, E. (2001). *Vitam. Horm.* **61**, 157–171.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Calder, R. B., Williams, R. S., Ramaswamy, G., Rock, C. O., Campbell, E., Unkles, S. E., Kinghorn, J. R. & Jackowski, S. (1999). *J. Biol. Chem.* **274**, 2014–2020.
- Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Engel, C. & Wierenga, R. (1996). *Curr. Opin. Struct. Biol.* **6**, 790–797.
- Datta, S., Krishna, R., Ganesh, N., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2003). *J. Bacteriol.* **185**, 4280–4284.
- Datta, S., Prabu, M. M., Vaze, M. B., Ganesh, N., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2000). *Nucleic Acids Res.* **28**, 4964–4973.
- French, G. S. & Wilson, K. S. (1978). *Acta Cryst.* **A34**, 517.
- Goulding, C. W. *et al.* (2002). *Curr. Drug. Targets Infect. Disord.* **2**, 121–141.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
- Leipe, D. D., Koonin, E. V. & Aravind, L. (2003). *J. Mol. Biol.* **333**, 781–815.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Rock, C. O., Calder, R. B., Karim, M. A. & Jackowski, S. (2000). *J. Biol. Chem.* **275**, 1377–1383.
- Rock, C. O., Park, H.-W. & Jackowski, S. (2003). *J. Bacteriol.* **185**, 3410–3415.
- Roy, S., Gupta, S., Das, S., Sekar, K., Chatterji, D. & Vijayan, M. (2004). *J. Mol. Biol.* **339**, 1103–1113.
- Saikrishnan, K., Jeyakanthan, J., Venkatesh, J., Acharya, N., Sekar, K., Varshney, U. & Vijayan, M. (2003). *J. Mol. Biol.* **331**, 385–393.
- Terwilliger, T. C. *et al.* (2003). *Tuberculosis*, **83**, 223–249.
- Yun, M., Park, C. G., Kim, J. Y., Rock, C. O., Jackowski, S. & Park, H.-W. (2000). *J. Biol. Chem.* **275**, 28093–28099.
- Zhou, B., Westaway, S. K., Levinson, B., Johnson, M. A., Gitschier, J. & Hayflick, S. J. (2001). *Nature Genet.* **28**, 345–349.