Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Invariance and variability in bacterial PanK: a study based on the crystal structure of *Mycobacterium tuberculosis* PanK

Satyabrata Das, Parimal Kumar, Vikrant Bhor, A. Surolia and M. Vijayan*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Correspondence e-mail: mv@mbu.iisc.ernet.in

Received 27 January 2006
Accepted 7 April 2006

PDB References: MtPanK, form I, room temperature, 2ges, r2gessf; low temperature, 2get, r2getsf; form II, room temperature, 2geu, r2geusf; low temperature, 2gev, r2gevsf.

Pantothenate kinase (PanK) is a ubiquitous and essential enzyme that catalyzes the first step of the universal coenzyme A biosynthetic pathway. In this step, pantothenate (vitamin B_5) is converted to 4′-phosphopantothenate, which subsequently forms coenzyme A in four enzymatic steps. The complex of this enzyme from *Mycobacterium tuberculosis* (MtPanK) with a derivative of the feedback inhibitor coenzyme A has been crystallized in two forms and its structure solved. The structure was refined in both forms using room-temperature and low-temperature X-ray data. In both forms, the MtPanK subunit has a mononucleotide-binding fold with a seven-stranded central β-sheet and helices on either side. However, there is a small though significant difference in subunit association between the two forms. The structure is also grossly similar to the enzyme from *Escherichia coli*. The active-site pocket and the dimeric interface are on two opposite sides of the PanK subunit. The enzymes from *M. tuberculosis* and *E. coli* exhibit several differences, particularly at the dimeric interface. On the other hand, the coenzyme A-binding region is almost entirely conserved. A delineation of the invariant and variable features of the PanK structure further indicates that the dimeric interface is very variable, while the coenzyme A-binding site is substantially invariant. A sequence alignment involving various bacterial PanKs is in agreement with this conclusion. The strong correlation between structural plasticity, evolutionary conservation and variability and function exhibited by the molecule could be important in the design of species-specific inhibitors of the enzyme.

1. Introduction

Pantothenate kinase (PanK) is an essential enzyme that is present in almost all living organisms (Leonardi *et al.*, 2005). It is a small-molecule kinase and performs ATP-mediated phosphorylation of the hydroxyl group of pantothenate (vitamin B_5) to form 4′-phosphopantothenate. This phosphorylation of pantothenate is the first step in the universal five-step coenzyme A (CoA) biosynthesis pathway (Begley *et al.*, 2001; Daugherty *et al.*, 2002; Kupke *et al.*, 2003; Genschel, 2004). The activity of PanK is feedback-regulated by CoA and its derivatives (Vallari *et al.*, 1987; Zhang *et al.*, 2005). It is a key regulatory enzyme in the CoA biosynthetic pathway, which is the primary mechanism by which bacteria maintain the intracellular concentration of CoA (Rock *et al.*, 2000,

2003). In the case of mammals, such as mouse and human, four PANK genes have been isolated: PANK1–PANK4 (Zhou *et al.*, 2001; Zhang *et al.*, 2006). The multiple isoforms of PanK coded by these four genes vary widely in length and exhibit tissue-specific expression. The sequences of all of them differ substantially from those of bacterial PanK (Cheek *et al.*, 2005).

CoA is an important cofactor and interacts with 4% of known enzymes with diverse folds and functions (Engel & Wierenga, 1996). However, in PanK CoA binds as a feedback inhibitor and hence its role is completely distinct from that in other proteins. Presently, the only known structure of PanK is that from *Escherichia coli*, EcPanK, in complex with CoA, with 5′-adenylimidodiphosphate, a non-hydrolysable analogue of ATP, and with ADP as well as pantothenate (Yun *et al.*, 2000; Ivey *et al.*, 2004). EcPanK is a homodimeric protein. Each subunit (MW 36 kDa) consists of 316 amino acids. It is a P-loop kinase and belongs to the mononucleotide-binding fold, with a seven-stranded central β-sheet flanked by helices on either side (Leipe *et al.*, 2003). Two long antiparallel helices are present in the dimerization interface. Kinetic studies show that the binding of ATP to EcPanK is highly cooperative and that CoA inhibits activity by competitive binding to the ATP site (Song & Jackowski, 1994). This cooperativity is presumably related to the dimeric nature of the enzyme. However, the structural basis of the propagation of the allosteric signal from one subunit to the other is not known. In the crystal structure of EcPanK, ATP and CoA interact with the enzyme 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. Further details of the mechanism of action have yet to be fully elucidated.

Mycobacterium tuberculosis (Mt), the causative agent of tuberculosis (TB), is among the best studied Gram-positive bacteria. The availability of the complete sequence of its genome (Cole *et al.*, 1998) has provided further impetus to the study of TB proteins. As a part of a global effort (Terwilliger, 2000; Vijayan, 2005) on the structure determination of such proteins, we have been pursuing a programme on the X-ray studies of proteins from *M. tuberculosis* and the related *M. smegmatis* (Datta *et al.*, 2000, 2003; Saikrishnan *et al.*, 2003; Roy *et al.*, 2004; Saikrishnan, Kalapala *et al.*, 2005; Saikrishnan, Manjunath *et al.*, 2005). As part of this programme, we report here the crystal structure of MtPanK in complex with a CoA derivative in two related crystal forms at two temperatures. This dimeric protein with a molecular weight of 35.7 kDa has a subunit containing 312 amino-acid residues. A comparison of the four structures among themselves and with the structure of EcPanK–CoA complex leads to the delineation of the structurally invariant and variable regions of this P-loop kinase molecule. This, coupled with an analysis of the known sequences of bacterial PanKs, provides interesting insights into the oligomerization and CoA binding of these proteins and their possible relevance to selective inhibitor design. In this context, it is interesting to note that PanK has been suggested to be essential for the growth of *M. tuberculosis* (Sassetti *et al.*, 2003).

2. Materials and methods

2.1. Protein purification, crystallization and data collection

The cloning, expression, purification and crystallization of MtPanK have been reported previously (Das *et al.*, 2005). Briefly, appropriate primers were selected in order to amplify the genomic DNA of MtPanK from *M. tuberculosis* H37Rv. The amplified gene with His_6 at the N-terminus was cloned into the expression vector pET-28a(+). The cloned gene was transformed into *E. coli* BL21 (DE3) cells. The cells were grown in LB medium and induced by IPTG (0.2 m*M*). A single-step purification of the recombinant protein was performed using an Ni–NTA metal affinity column. The purified protein was dialyzed into crystallization buffer consisting of 0.1 *M* Tris–HCl, 0.15 *M* NaCl, 5%(*v*/*v*) glycerol and 0.001 *M* β-mercaptoethanol pH 8.0. Crystals suitable for diffraction study were grown by the hanging-drop vapour-diffusion method. Crystallization droplets were prepared by mixing 3 µl protein solution (6 mg ml^{-1}) 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 equilibrated against 400 µl of the same precipitant solution. Two crystal forms grew under the same condition. An in-house MAR 345 image-plate system mounted on a Rigaku Ultrax-18 rotating-anode X-ray generator was used to record screenless oscillation images with an oscillation angle of 1°. Low-temperature (100 K) data sets were collected by flash-freezing the crystals in a liquid-nitrogen stream produced by an Oxford Cryosystem. Initially, 20% glycerol was added to the crystallization droplet; the crystals were soaked for 2–3 h and then transferred into 30% glycerol prior to flash-freezing for data collection. The recorded images were processed and scaled using *DENZO* and *SCALEPACK* from the *HKL* package (Otwinowski & Minor, 1997) and the intensity data were truncated to amplitudes using *TRUNCATE* from the *CCP*4 suite (French & Wilson, 1978; Collaborative Computational Project, Number 4, 1994).

2.2. Structure solution and refinement

The structures of both crystal forms were solved by molecular replacement (Das *et al.*, 2005) using *AMoRe* (Navaza, 1994) with the EcPanK structure (PDB code 1esm) as the model. Refinement was performed using *CNS* v.1.1 (Brünger *et al.*, 1998) and model building was carried out using the interactive graphics program *FRODO* (Jones, 1978). In all the structures, electron density for CoA was clearly visible in both the $2F_o - F_c$ and $F_o - F_c$ maps when the *R* factor was around 0.30 and the CoA was incorporated into the model. The topology and parameter file for CoA was generated using the *PRODRG* server (Schüttelkopf & van Aalten, 2004). After further refinement, water molecules were located at positions with peak heights 0.8σ and 2.5σ in the $2F_o - F_c$ and $F_o - F_c$ maps, respectively. Several glycerol molecules were also located in the low-temperature structures. Careful inspection of the electron-density maps showed extra density beyond the free sulfhydryl group of CoA and also Cys173, which were

modelled as adducts formed with the β-mercaptoethanol present in the protein buffer. The topology and parameter files for the *S*-(thioethylhydroxy)coenzyme A were then modified by the *PRODRG* server and those for *S*,*S*-(2-hydroxyethyl)-thiocysteine was obtained from the *HIC-UP* server (Kleywegt & Jones, 1998). The models were then refined by the slow-annealing protocol of *CNS* followed by *B*-factor refinement. Several simulated-annealing omit maps and a 1/8th omit map (Vijayan, 1980) were calculated to remove model bias.

2.3. Analysis of sequence and structure and validation

The structures were validated using the program *PROCHECK* (Laskowski *et al.*, 1993). Contact distances were calculated using the program *CONTACT* from *CCP*4 (Collaborative Computational Project, Number 4, 1994). Structural alignments were carried out using *ALIGN* (Cohen, 1997). Buried surface area was calculated using *NACCESS* (Hubbard & Thornton, 1993). The program *ESCET* (*Error-inclusive Structure Comparison and Evaluation Tool*) used in this work categorizes the subjected molecule into conformationally variable and invariant regions by automated analysis based on genetic algorithm of all the plausible pairs of error-scaled difference matrices (EDD) of an ensemble of conformers (Schneider, 2002). Sequence alignment of the 29 bacterial PanK sequences was performed using program *CLUSTALW* (Thompson *et al.*, 1994). The *PHYLIP* package (v.3.6a3) was used for phylogenetic analysis (Felsenstein, 1989). The aligned sequences from the *CLUSTALW* were manually adjusted.

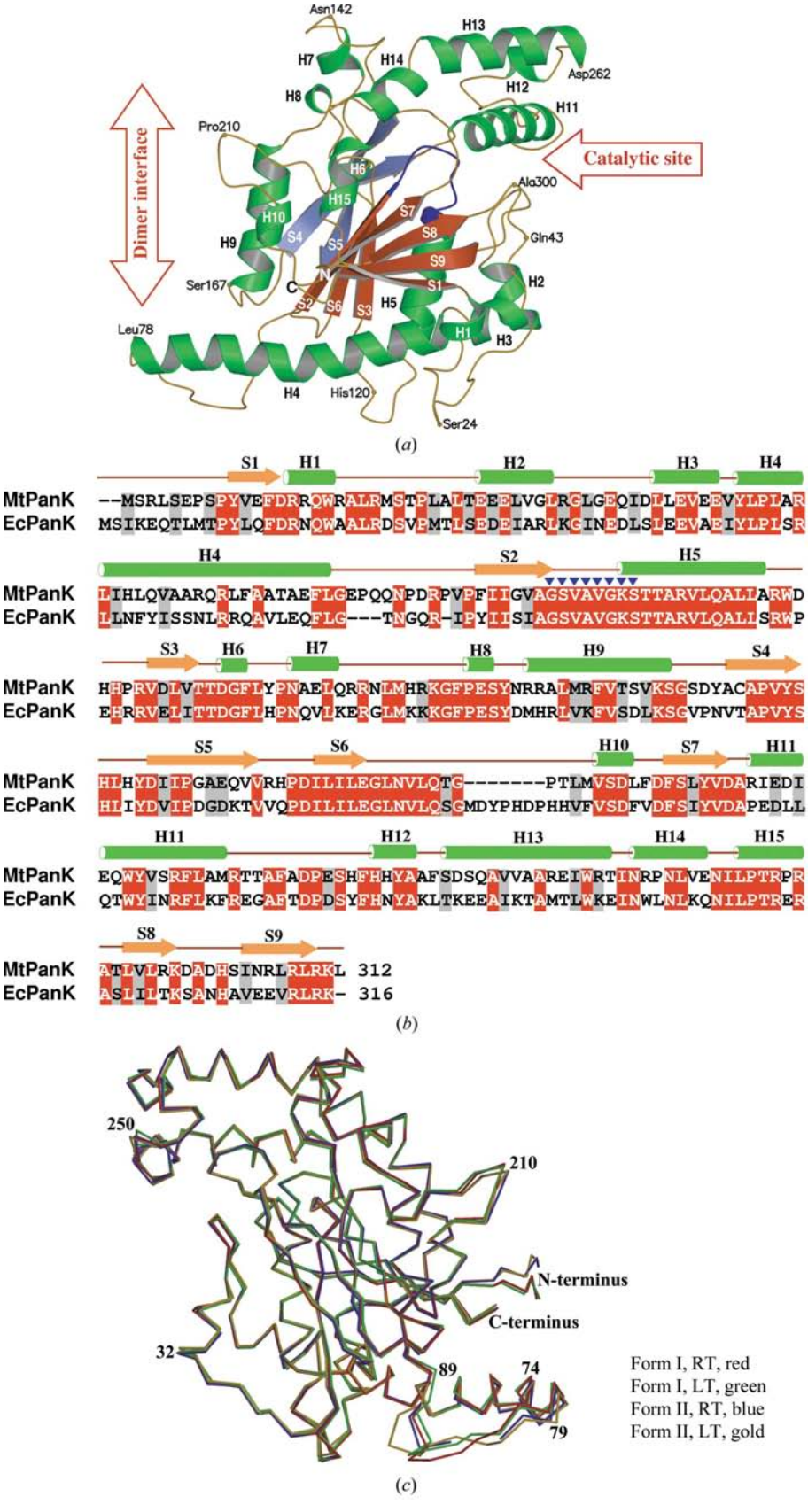

Figure 1
(*a*) Ribbon representation of the MtPanK subunit. β-Strands are labelled S1–S9 and helices are labelled H1–H15 (green). The seven-stranded central β-sheet (S1–S3 and S6–S9) is shown in red and the two-stranded β-sheet (S4 and S5) in light blue. The phosphate-binding loop is in deep blue. S6 constitutes the Walker B motif. The N-terminus and C-terminus are marked as N and C and several C^{α} atoms are labelled. (*b*) The sequences of MtPanK and EcPanK aligned by *CLUSTALW* (v.1.81). Helices (H1–H15) and strands (S1–S9) of MtPanK are depicted as cylinders (green) and arrows (orange). P-loop residues are marked as blue-coloured triangles. (*c*) Superposition of subunits in the four structures of MtPanK.

Table 1
X-ray crystal data, refinement and model statistics.

Values in parentheses are for the highest resolution shell. RT, room temperature (293 K); LT, low temperature (100 K).

	Form I		Form II	
	RT	LT	RT	LT
Crystal data				
Space group	$P3_121$	$P3_121$	$P3_121$	$P3_121$
Unit-cell parameters (Å)				
a	78.3	76.97	107.63	104.39
c	115.45	113.65	89.85	90.59
V_M (Å^3 Da^{-1})	2.8	2.7	4.1	3.9
Solvent content (%)	55.6	53.3	69.8	68.2
No. of subunits per ASU	1	1	1	1
Resolution limits (Å)	20.0–2.4	30.0–2.35	20.0–2.9	30.0–2.35
Highest shell (Å)	2.49–2.4	2.43–2.35	3.0–2.9	2.43–2.35
Total No. of measured reflections	95126	165465	62923	362603
Total No. of unique reflections	16374	16679	13523	24164
Completeness (%)	99.2 (97.1)	99.6 (99.9)	99.0 (99.7)	99.8 (99.8)
R_{merge}† (%)	8.3 (52.6)	5.9 (56.2)	8.8 (53.6)	6.4 (50.3)
Average $I/\sigma(I)$	13.5 (2.5)	22.7 (3.7)	14.0 (3.3)	25.6 (4.1)
Multiplicity	5.8 (3.0)	9.9 (6.4)	4.7 (4.7)	15.0 (10.0)
Refinement and model statistics				
R factor (%)	18.8	22.7	18.3	21.3
R_{free}‡ (%)	23.2	27.0	22.1	23.0
Total No. of atoms per ASU				
Protein	2477	2483	2448	2482
Water	178	176	86	286
CoA	52	52	52	52
Glycerol	—	12	—	36
Tris	—	—	—	8
R.m.s. deviation from ideal				
Bond lengths (Å)	0.006	0.007	0.007	0.006
Bond angles (°)	1.4	1.3	1.4	1.4
Dihedral angles (°)	23.6	23.7	25.0	23.8
Improper angles (°)	0.95	0.90	0.89	0.98
Ramachandran plot statistics (% of residues)				
Core regions	90.9	87.2	90.0	90.9
Allowed regions	6.9	11.3	9.3	6.6
Generously allowed regions	2.2	1.5	0.7	2.6
Disallowed regions	0.0	0.0	0.0	0.0

† $R_{merge} = \sum |I(k) - \langle I \rangle| / \sum |I(k)|$, where $I(k)$ is the kth intensity measurement of a reflection, $\langle I \rangle$ is the average intensity value of that reflection and the summation is over all measurements. ‡ 5% of reflections were used for the R_{free} calculations.

Bootstrapping was performed using *SEQBOOT*. Distance matrices were generated using *PROTDIST* from the *PHYLIP* package and the significance of the neighbour-joining tree thus obtained were evaluated using the program *CONSENSE*. The final tree was constructed using the program *DRAWGRAM*.

The figures were prepared using *BOBSCRIPT* (Esnouf, 1997), *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Bacon, 1997).

3. Results and discussion

3.1. Overall structure of MtPanK

The structure of MtPanK (Fig. 1*a*) has been determined in two forms (form I and form II) and refined against data collected at room temperature (RT) and low temperature (LT) in each case (Table 1). Both of the LT data sets and the RT data set from form I contained data to a resolution of 2.4 Å or better. The resolution of the RT structure of form II is only 2.9 Å. The first four residues in the 312-residue chain are disordered in all the structures. In addition, residues 82–84 had no electron density in the RT structure of form II. In all the structures Cys173 has extra density beyond the SH group which has been interpreted as resulting from the formation of an adduct with β-mercaptoethanol in the crystallization medium that converts free cysteine to *S*,*S*-(2-hydroxyethyl)thiocysteine (Fig. 2*a*). This type of adduct formation is well known in the literature and more than 50 PDB entries containing such adducts have been deposited in the PDB (Berman *et al.*, 2000). Each structure has a PanK subunit in the asymmetric unit. The physiological dimer is generated by a crystallographic twofold axis.

As in the case of EcPanK, the subunit in all four structures has a mononucleotide-binding fold with a central seven-stranded β-sheet made up of strands S1–S3 and S6–S9 (Figs. 1*a* and 1*b*). Six of the strands are parallel to one another, while S9 is antiparallel to the rest. There is another less extensive antiparallel β-sheet made up of strands S4 and S5. The central sheet is flanked by helices on both sides. There are several more helices in the structure, some with distinctive roles that are discussed later. In fact, twice as many residues are in helices as are in sheets. The helices and strands are interconnected by loops, some highly elaborate, which make up a third of the structure. The P-loop, involved in nucleotide binding, resides in the peptide stretch connecting S2 and H5 (Fig. 1*b*). The strand S6 constitutes the Walker B motif and is adjacent to S2 in the central β-sheet.

Pairs of subunits from the four structures superimpose with r.m.s. deviations in C^α positions ranging from 0.4 to 0.85 Å when all 308 (or 305 when one of the structure is form II RT) residues are included in the calculations. When highly deviating residues are excluded, as normally performed when using *ALIGN* (Cohen, 1997), the range decreases to 0.28–0.45 Å, with the number of residues superposed ranging from 286 to 302. As illustrated in Fig. 1(*c*), the deviations are localized in a few regions. The stretch ranging from residues 72 to 88, which encompasses the C-terminal quarter of H4 and much of the loop that follows it, exhibit deviations to different degrees depending upon the pairs of structures considered. Strikingly large deviations are observed in every pair in the loop. The N-terminal stretch involving residues 5–8 is also prone to structural deviations. The other residues that show substantial deviations in one or more pairs of residues include 35–36, 171, 244 and 209–211. These deviations provide a

measure of the inherent flexibility of MtPanK and, as shown later, are of structural and biological significance.

The common mode of dimerization in the four structures is illustrated in Fig. 3(*a*). Helix H4 and the residues that immediately follow it play a major role in stabilizing the dimer; the helices from the two subunits align in an antiparallel manner in a direction nearly perpendicular to the molecular dyad and interact with each other. In fact, much of H4 (residues 59–78) and its partner from the other subunit form a structure akin to an antiparallel two-helix coiled coil with heptad repeat (Walshaw & Woolfson, 2001). The axes of the helices make an angle of 171° and they are 8.3 Å apart. These values are well within the range of those observed in other antiparallel two-stranded coiled coils. The other regions involved in dimerization encompass the C-terminal half of H9, many residues, particularly 208–212, in the loop between S6 and H10, H10 and H15 and residues at the N- and C-termini. Interestingly, there is considerable commonality in the residues which exhibit deviations among the four structures and those involved in dimerization.

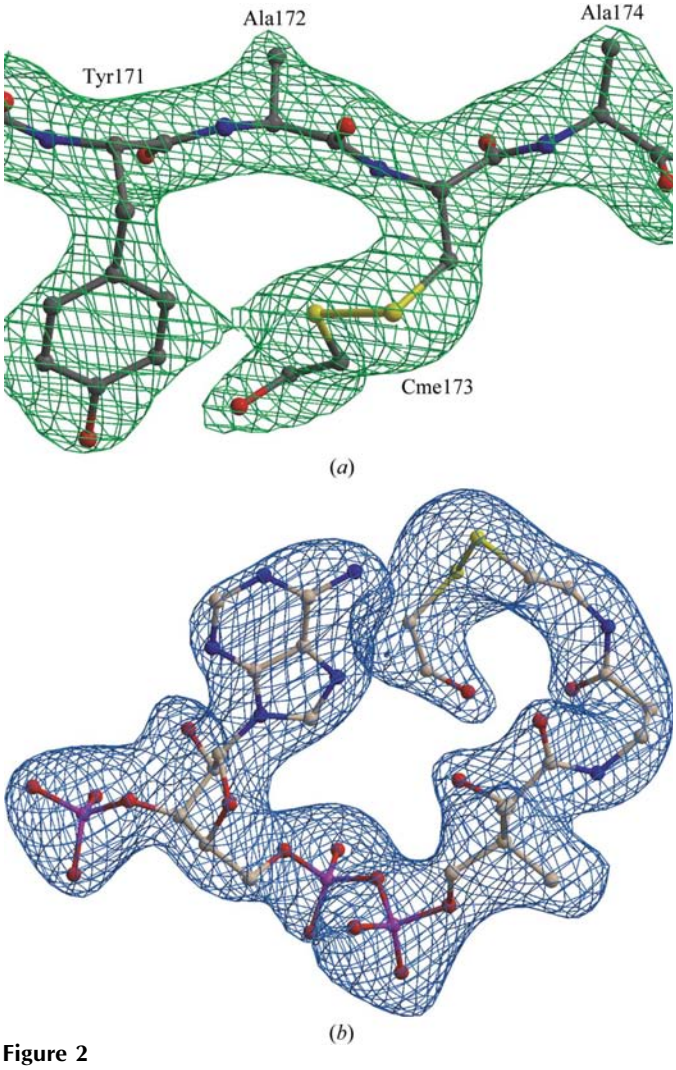

Figure 2
Simulated-annealing omit map with Fourier coefficients $F_o - F_c$ (3σ) of the adduct (*a*) *S*,*S*-(2-hydroxyethyl)thiocysteine and (*b*) *S*-(thioethyl-hydroxy)coenzymeA.

The residues involved in dimerization are almost the same in the four structures. However, the subunits within the dimer have slightly different mutual orientations in the two crystal forms. As illustrated in Fig. 3(*b*), the difference can be explained in terms of a rotation of one subunit with respect to the other about an axis perpendicular to the molecular crystallographic dyad. This axis is roughly parallel to the H4 helices from the two subunits. The angle of rotation is 5.8° between the two room-temperature forms. It is 6.9° when the two low-temperature forms are compared.

The surface area buried on dimerization in the four structures, along with that in the two dimers in the EcPanK structures, is given in Table 2. More than two-thirds of the buried area is non-polar in nature. A salt bridge between the side chains of Glu6 of one subunit and Arg160 of the other exists in both forms. The other intersubunit hydrogen bonds that occur in both the forms are Arg58 NH1···Phe77 O,

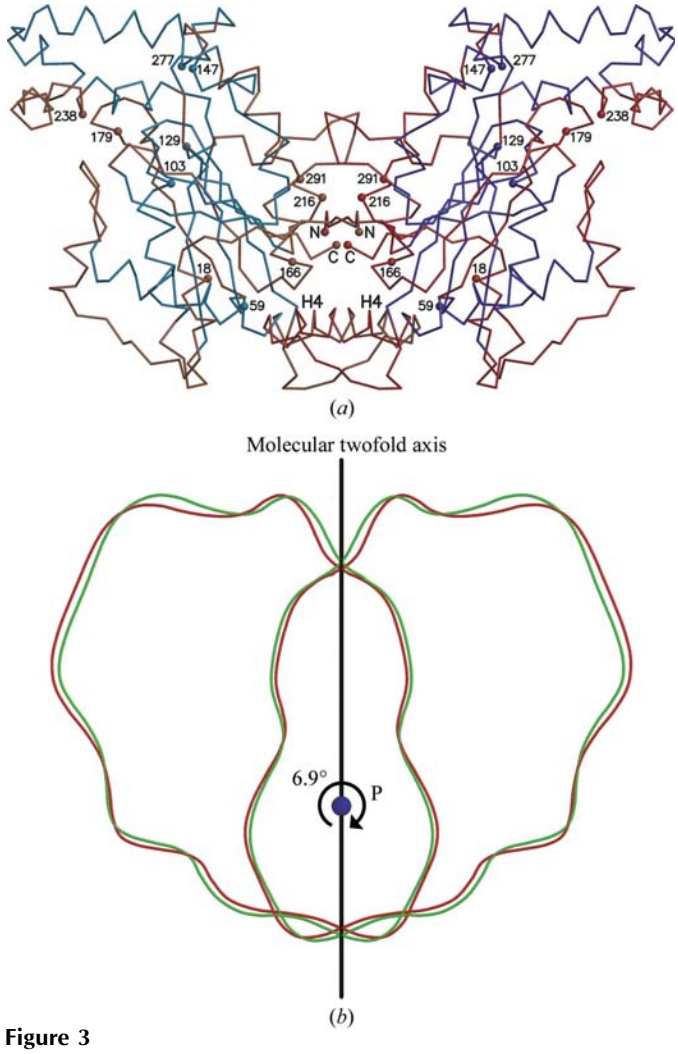

Figure 3
(*a*) C^{α} trace of the homodimeric MtPanK molecule. The invariant and the variable regions are shown in blue and red, respectively. See text for details. (*b*) Schematic representation of the superposition of MtPanK dimers in crystal form I (red) and form II (green), respectively. Their twofold axes are aligned. P represents the axis about which one subunit has to be rotated for perfect superposition. This angle of rotation in the case of the low-temperature structures is indicated. See text for details.

Arg58 NH2···Glu76 O, Gln63 NE2···Thr74 OG1 and Arg310 NH2···Asp216 O. In all four structures Gly209 O of one subunit is involved in at least one hydrogen bond to the guanidyl group of Arg291 of the other subunit. There are a few interactions which occur in only one of the two forms. Prominent among them are the interactions of the guanidyl group of Arg291 of one subunit with Asp216 and a couple of other residues in the second subunit. They occur only in form I (Fig. 4*a*). The side chain of Arg291 has a different conformation, which in form II structures enables it to form an internal hydrogen bond to a carbonyl O atom of the same subunit (Fig. 4*b*). Interestingly, in the absence of intermolecular interactions between Arg291 and Asp216, part of the 202–213 loop, particularly residues 204–210, in one subunit and the C-terminus of H15 (289–291) in the other move close to each other, resulting in a larger burial of non-polar surface in addition to the possible formation of a Thr208 OG1···Thr208 OG1 hydrogen bond. As can be seen from Table 2, the non-polar surface area buried on dimerization is larger by more than 180 Å^2 in form II. Much of the contributions to it come from the 208–217 and 287–291 stretches. Thus, it appears that the loss of hydrogen bonds involving Arg291 in form II is compensated by non-bonded and hydrophobic interactions.

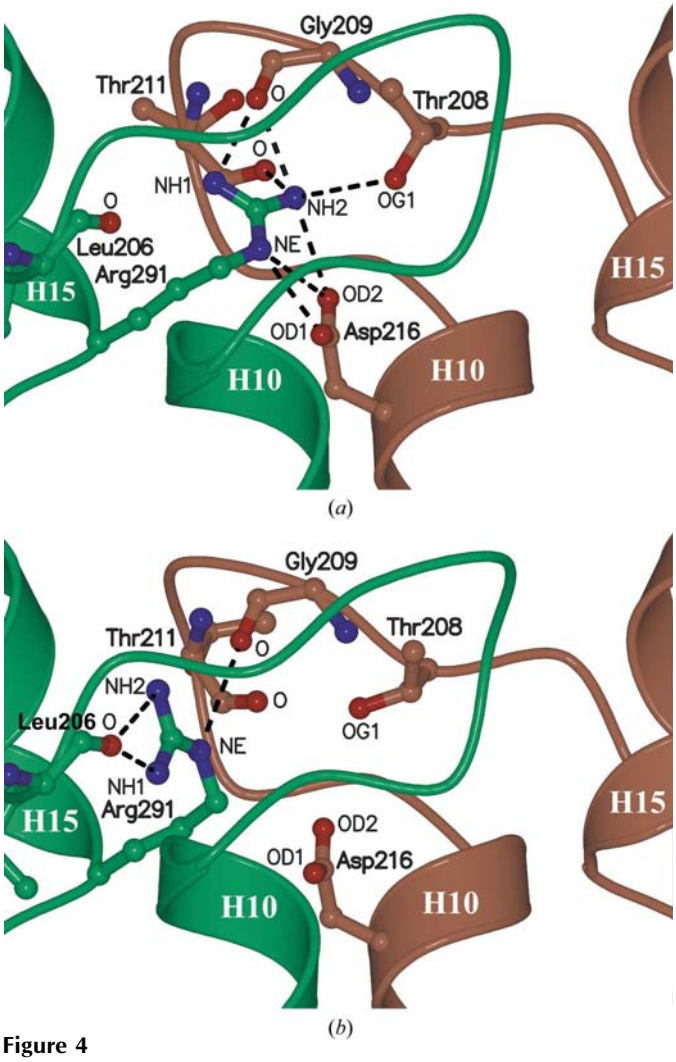

Figure 4
The side-chain orientations of Arg291 and its interactions in (*a*) form I and (*b*) form II.

3.2. Structural comparison with EcPanK

The crystal structure of the EcPanK–CoA complex has two nearly identical dimers in the asymmetric unit. However, the two subunits in each dimer exhibit differences in structure, as illustrated in Fig. 5(*a*). The difference is most conspicuous in the region involving the C-terminal region of H4 and the subsequent loop (residues 70–85). This region has four deletions in the EcPanK sequence (Fig. 1*b*). Another such region involves the loop between S6 and H10. This loop is characterized by a seven-residue insertion in EcPanK. Residues 8

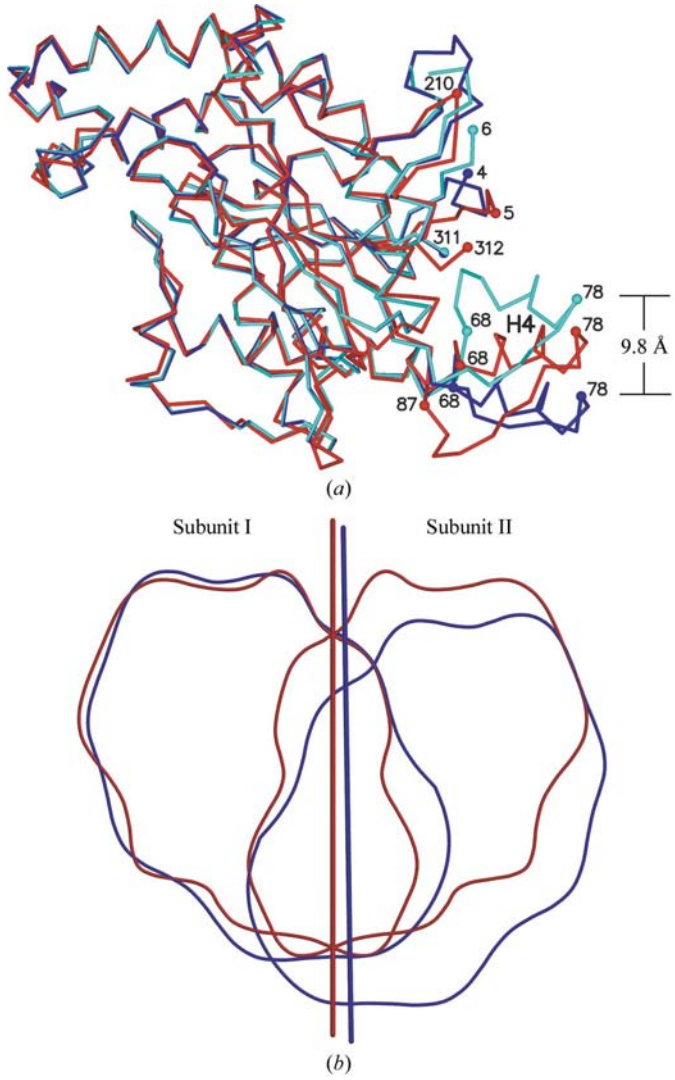

Figure 5
(*a*) The C$^\alpha$-atom superposition of one subunit of the MtPanK dimer (red) and the two crystallographically independent subunits of EcPanK, subunit *A* (blue) and subunit *B* (cyan). (*b*) Schematic representation of the relative orientation of the two subunits in the MtPanK and EcPanK dimers. The dimeric molecules are overlayed based on the alignment of subunit I. The molecular twofold axes are shown in the same colour as that used to represent the respective molecules.

Table 2
Accessible surface area buried on dimerization in MtPanK and EcPanK.

The PDB code of EcPanK is 1esm and those of MtPanK are 2ges (form I, RT), 2get (form I, LT), 2geu (form II, RT) and 2gev (form II, LT).

	MtPanK					
	Form I		Form II		EcPanK	
	RT	LT	RT	LT	Dimer I	Dimer II
Buried surface area (Å^2)	4142.5	4001.5	4130.2	3958.6	3309.3	3306.6
Non-polar (% value)	2836.6 (68.5)	2782.6 (69.5)	3027.7 (73.3)	2951.3 (74.5)	2502.9 (75.6)	2372.1 (71.7)

and 9 at the N-terminus also show substantial deviations. Interestingly, in the 70–85 region, the stretches in subunits *A* and *B* in EcPanK lie on either side of that in MtPanK (Fig. 5*a*). This region is involved in dimerization.

Although the structures of the two subunits are essentially the same except for the differences mentioned above, the EcPanK–CoA complex, unlike the corresponding MtPanK complex, does not form perfectly twofold symmetric dimers. In each dimer, one subunit is related to the other by a rotation of nearly 180° and a translation of about 5.5 Å along the rotation axis. Consequently, as illustrated in Fig. 5(*b*), the structure of a dimer of the EcPanK complex is somewhat different from that of the dimer of the MtPanK complex.

The intersubunit interactions in the EcPanK complex are less extensive than those in the MtPanK complex. The surface area buried in the former is lower by about 650–830 Å^2 than in the latter (Table 2). The numbers of intersubunit hydrogen bonds in the two complexes are nearly the same. Thus, the dimeric structure of MtPanK complex can be expected to be somewhat stabler than that of the EcPanK complex. However, interestingly, the regions of the molecule involved in dimerization are nearly the same in the two complexes. This is also reflected in the hydrogen-bonding pattern. The details of hydrogen bonding differ, but the regions connected by hydrogen bonds exhibit similarities. Interestingly, peptide stretches involved in deletions and insertions when the sequences of Ec and MtPanK are compared are involved in dimerization. These stretches exhibit considerable structural variation between the four structures of the MtPanK complex, between subunits *A* and *B* of the EcPanK complex and also between MtPanK and EcPanK. The other regions involved in the dimerization also exhibit substantial structural variability. Thus, the dimerization of PanK appears to involve structurally variable features.

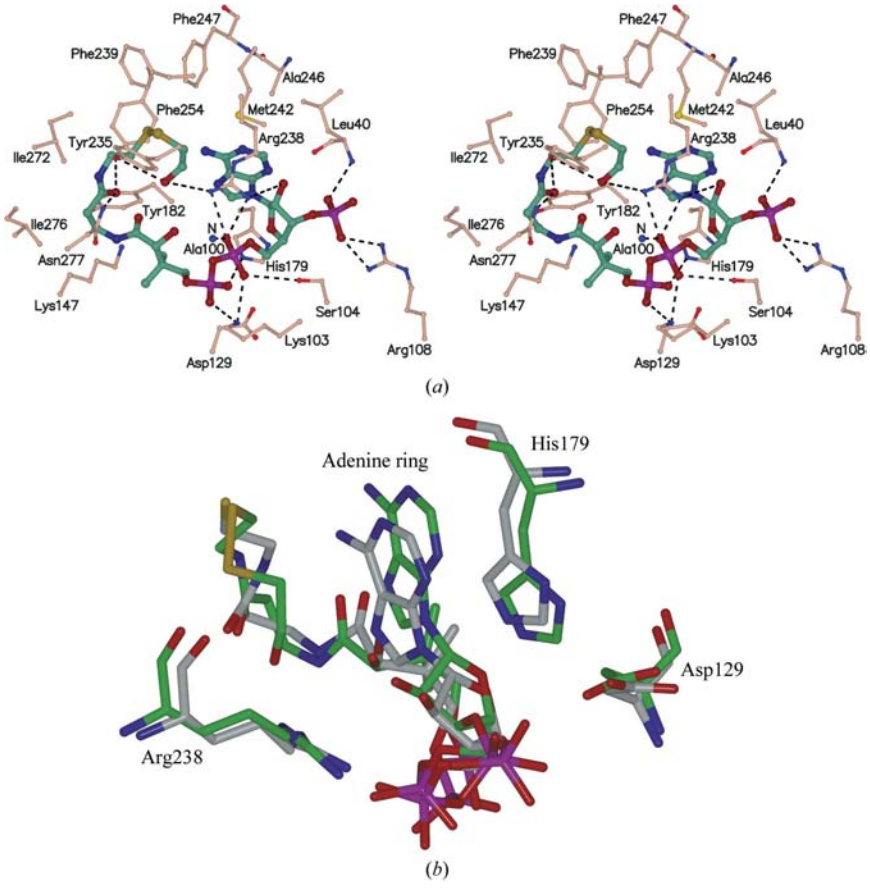

Figure 6
(*a*) Stereoview of MtPanK–CoA interactions. The CoA derivative and the binding site of the enzyme are shown in cyan and brown, respectively. Dotted lines represent possible hydrogen bonds. (*b*) Positions of the adenine ring in the CoA with respect to the three catalytic residues Asp129, His179 and Arg238 as observed in MtPanK–CoA (green) and EcPanK–CoA (grey).

3.3. PanK–CoA interactions

CoA-binding proteins have very diverse folds and functions (Engel & Wierenga, 1996). However, PanK is the only known example where a protein binds CoA as an inhibitor. All the MtPanK structures have well defined electron density corresponding to CoA. In all of them, clear density extends well beyond the terminal SH group. Several possibilities were explored to account for this extra density. It appeared reasonable to assume that the β-mercaptoethanol in the crystallization medium reacted with CoA to yield *S*-(thioethylhydroxy)coenzyme A (Fig. 2*b*). CoA in the MtPanK complex has a bent conformation as in the EcPanK complex (Yun *et al.*, 2000).

The CoA-binding region in MtPanK has almost the same structure as that in EcPanK. The accessible surface area of 29 residues in the enzyme is affected when

CoA binds to MtPanK. EcPanK has identical residues at 26 out of these 29 positions. The residues are conserved in the remaining three positions (Leu40/Ile40, Met242/Phe242 and Ile272/Leu272). CoA is almost entirely buried within the protein; nearly 90% of the accessible surface area of the molecule is buried on complexation. The adenine moiety points towards the protein and 3′-phosphate is partly exposed, a feature that is also observed in other CoA complexes (Engel & Wierenga, 1996).

In consonance with the chemical nature of the ligand, the CoA-binding site involves hydrophobic and hydrophilic surfaces (Fig. 6*a*). Tyr235, Phe239 and Met242 in H11 and Ala246, Phe247 and Phe254 of the loop that follows are among the main contributors to the hydrophobic surface. The other contributors include Leu40 and Tyr182, which belong to loops, Ile272 of H13 and Ile276, which belongs to the short loop that follows H13. The hydrophobic surface is in contact with the pantothenate moiety and the base. The carbonyl group of the pantothenate moiety is hydrogen bonded to the amide group of Asn277 and the hydroxyl group of Tyr235. The P-loop (residues 97–104) is understandably involved in interactions with the phosphates that link the nucleoside and the pantothenate. Lys103 interacts with two adjacent phosphate groups, while Ala100 N and Ser104 OG are hydrogen bonded to one of them. The same phosphate group forms a salt bridge with the guanidyl group of Arg238. His179 NE2 forms a hydrogen bond with the other phosphate group. The free phosphate group attached to the sugar also forms a hydrogen bond, now with the guanidyl group of Arg108. The other hydrogen bond that this phosphate group makes is with Leu40 N. Glu201 of the Walker B motif is involved in an intramolecular salt bridge with Lys103 NZ of the P-loop, which is absent in the EcPanK–CoA complex. In EcPanK, the adenine ring of the CoA is stacked between residues Phe242 and His179. In MtPanK the residue at 242 is Met and as a consequence there is a change in the position of the adenine ring with respect to the three proposed catalytic residues (Asp129, His179 and Arg238) when the Mt and Ec complexes are compared (Fig. 6*b*).

A major difference between the MtPanK and the EcPanK complexes is in the size and composition of the ligand. In EcPanK the ligand is CoA, while it is *S*-(thioethylhydroxy)-CoA in MtPanK. The additional thioethylhydroxy group does not cause any appreciable change in the ligand-binding site. This prompted an examination of the effect of the binding of acetyl-CoA to the enzyme. The reduced affinity of this ligand compared with that of CoA itself for EcPanK has been attributed to possible steric clashes of the acetyl group with the binding site (Yun *et al.*, 2000). To test this suggestion, the acetyl group was modelled into the EcPanK–CoA complex and the group was rotated about the possible torsion angles. This led in a straightforward manner to several conformations not involving any steric clash. In fact, one of them even has additional hydrogen bonding involving the carbonyl group of the acetyl moiety and the OH group of Tyr235. A similar exercise was carried out in the case of the MtPanK complex, which led to the same result. Thus, the reduced affinity of acetyl-CoA compared with that of CoA for the enzyme does not appear to be caused by steric clashes of the acetyl group with the binding site. Favourable interactions between the free SH group in CoA and the surrounding aromatic residues and an intramolecular hydrogen bond with the amino group of the adenine base are involved in the EcPanK complex. It is a moot question whether the modifications in these interactions caused by acetylation would lead to the reduced affinity.

3.4. Invariant and variable structural features

The four copies of the subunit of the MtPanK complex from two crystal forms, each analyzed at two different temperatures, and the four copies of the subunit of EcPanK complex belonging to two crystallographically independent dimers provide a database for exploring the invariant and variable structural features of the protomer. This was attempted using the program *ESCET*, which categorizes the molecule into conformationally invariant and variable regions by automated analysis of pairs of error-scaled difference distances of an ensemble of conformers. In the present calculations, the variable parameter (multiple of σ) was so chosen as to have roughly equal number of residues in the invariant and variable regions (Schneider, 2002).

The above calculations indicated the following residues to be in relatively conformationally variable regions: 5–44, 65–88, 116–120, 144–146, 162–194, 204–219, 227–228, 238–252, 281–291, 300–301 and 310–312. This is illustrated in Fig. 3(*a*). Significantly, 80% of the residues involved in dimer formation are in structurally variable regions. On the contrary, nearly 60% of the residues involved in CoA binding are in an invariant region; there are, however, a substantial number of residues (about 40%) in the variable regions. The whole of the P-loop is conformationally invariant, as is Arg108, which interacts with the 3′-phosphate. The only three residues which interact with phosphate groups but are in flexible regions are Leu40, His179 and Arg238. The last two of these are also catalytic residues. Thus, the phosphate-binding region of the binding site shows overwhelming conformational invariance. On the other hand, the residues which interact with the base and the pantothenate moiety, including the hydrophobic patch, belong to variable regions. The structures of the complexes of EcPanK indicate that the ATP- and CoA-binding sites partially overlap, particularly the P-loop region, which is conformationally invariant in EcPanK and MtPanK. The residues involved in interactions with the base and the sugar of ATP exhibit substantial variability. The central β-sheet is substantially conformationally invariant. Six strands in the sheet and the helices in contact with them constitute a central invariant block. Not surprisingly, most of the residues in much of the outlying regions exhibit conformational variability.

3.5. Analysis of the bacterial PanK sequences

Bacterial PanK sequences were retrieved from the SWISS-PROT/TREMBL database and a data set of 29 sequences with less than 90% similarity was constructed. The length of the sequences varies from 300 to 330 residues and the pairwise

sequence identity varies from 38 to 80%. Multiple sequence alignment using *CLUSTALW* and subsequent phylogenetic analysis employing *PHYLIP* showed that the sequences cluster into four major groups (Fig. 7*a*): eubacteria (ten sequences), high-GC Gram-positive (four sequences), α-proteobacteria (five sequences) and γ-proteobacteria (ten sequences). Thus, the clustering of PanK sequences follows the taxonomic classification. Furthermore, the conservation scores for each residue position in the 29 sequences were calculated using the program *CONSURF* v.3.0 (Glaser *et al.*, 2003) and were then mapped onto the three-dimensional structure of MtPanK (Fig. 7*b*). The most conserved residues, represented in maroon, are clustered around the active-site region. The subunit interface in the dimer is poorly conserved.

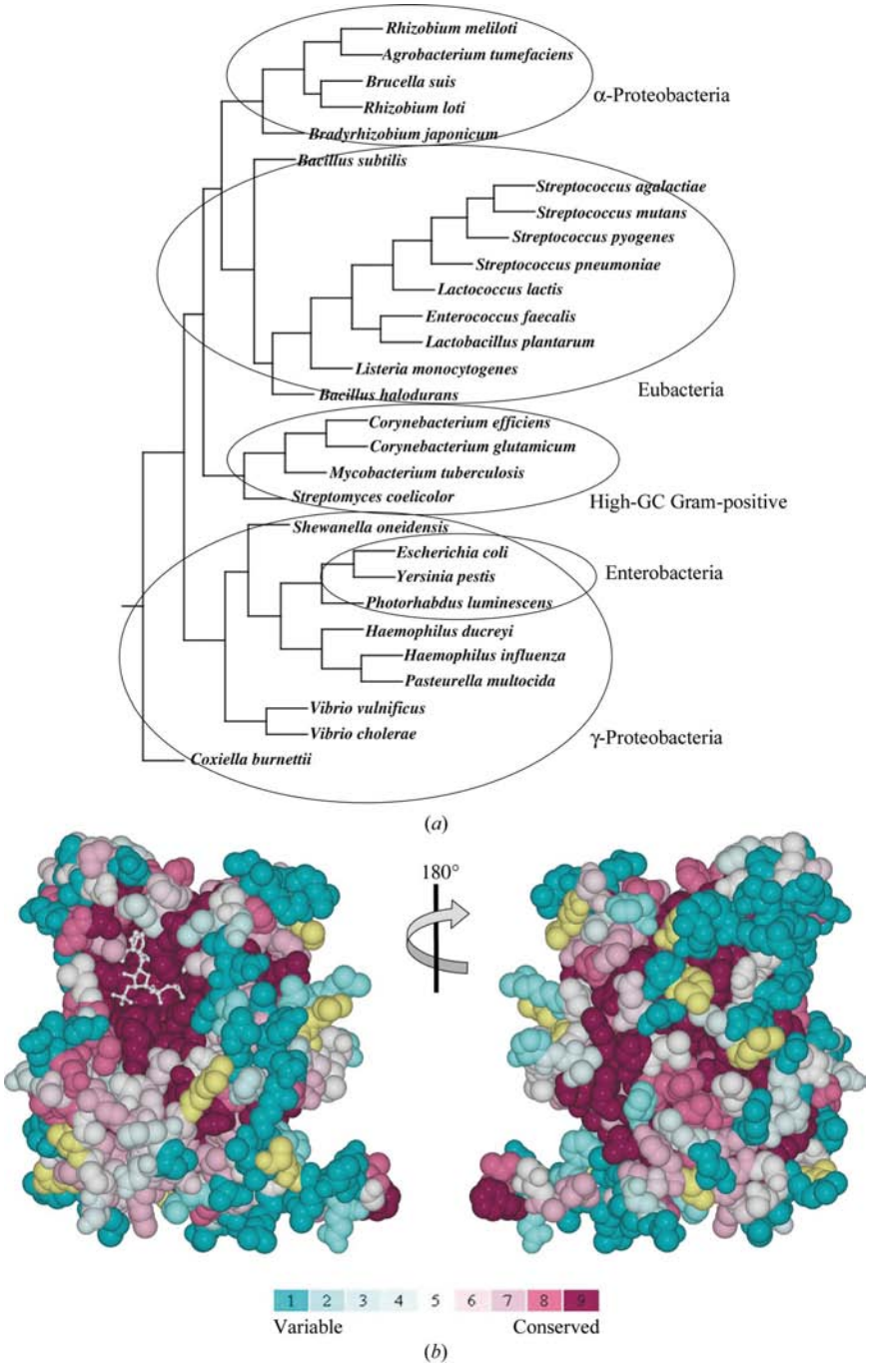

Figure 7
(*a*) Phylogenetic tree of the 29 bacterial pantothenate kinases. See text for details. (*b*) The evolutionary conservation score of the amino-acid positions in the 29 bacterial PanKs, generated by the *CONSURF* server and mapped onto the structure of MtPanK. The two figures are related by a rotation of 180° about the vertical axis. The right face of the figure on the left and left face of that on the right roughly correspond to the intersubunit interface. Bound CoA is shown in white in ball-and-stick representation.

There are a total of 53 identical residues in the 29 sequences. Of these, 39 belong to the conformationally invariant regions estimated using the MtPanK and EcPanK structures, thus indicating a broad correlation between structural invariance and evolutionary conservation. There are 37 residues whose accessible surface areas are affected by CoA and ATP binding in MtPanK and EcPanK. 20 of them are identical in all sequences. Another dozen identical residues also appear to be related in one way or other to nucleotide binding (residues in the P-loop whose accessible surface areas are not affected by nucleotide binding, residues which flank a catalytic residue *etc.*). These residues form part of a spatially contiguous cluster of 42 identical residues. The remaining residues in this cluster are probably important in maintaining the structural integrity of the region. Thus, a cluster comprising 80% of the residues identical in all the sequences constitutes the nucleotide-binding region.

There are about 50 residues whose accessible surface areas are affected by dimerization. Of these, only five, namely Trp18, Leu59, Lys166, Asp216 and Arg291, are identical in all the sequences. Asp216 and Arg291 are involved in an intersubunit salt bridge. Lys166 forms an intrasubunit salt bridge with Asp219 which is identical in all the sequences. The regions of the sequences which exhibit maximum variation in length (through insertions and deletions) are the N-terminal stretch, the loop between H4 and S2 and that between S6 and H10. All these regions are involved in the intersubunit interface. Thus, the regions involved in dimerization exhibit substantial sequence variation. The very few identical residues in the region appear to have specific structural roles.

Residues which are not identical, but are highly conserved, include Phe/Tyr77, Arg/Lys289 and Arg/Lys310, which deserve special mention. Arg289 forms an intrasubunit salt bridge with Asp225, while Arg310

forms an intersubunit hydrogen bond with Asp216. Both the Asp residues are identical in all the sequences. Phe/Tyr77 is situated at one end of the H4 helix and probably plays an important role in dimer formation. It exhibits the highest change in accessible surface area (more than 310 Å^2, most of which is non-polar) on dimerization in MtPanK and EcPanK. In this context, the C-terminal residue at position 312 in MtPanK also merits special mention. It is Leu or Ile in α-proteobacteria and high-GC Gram-positive bacteria, while the sequence only extends to position 311 in γ-proteobacteria. In MtPanK, an accessible surface area of about 240 Å^2 (more than 200 Å^2 non-polar) of this residue is buried on dimerization. The residue does not exist in EcPanK. Thus, this residue appears to contribute substantially to the higher stability of MtPanK compared with EcPanK.

4. Concluding remarks

The CoA-biosynthesis pathway, of which PanK is the first enzyme, is essential for the survival of the organism. The substantial differences between the human enzymes and the bacterial enzyme enhances its value as a possible drug target. The present study based on MtPanK and involving structural comparison with EcPanK and an analysis of the known sequences of bacterial PanKs leads to a picture of the molecule exhibiting strong correlation between structural plasticity, evolutionary conservation and variability and function. The core of the molecule made up of the nucleotide-binding region and the scaffold that supports it is substantially conformationally invariant and evolutionarily conserved. This is particularly true of the subregion which interacts with phosphates. The small differences in regions which interact with the rest of the nucleotides could serve to differentiate between bacterial PanKs in inhibitor design. Unlike the nucleotide-binding region, the intersubunit interface in the dimeric molecule exhibits substantial variability in structure and sequence. This interface, which could well be involved in the cooperativity of the action of the enzyme, could also be used to differentiate between bacterial PanKs and might turn out to be important in inhibitor design.

The intensity data sets were recorded using the in-house 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 carried out at the Supercomputer Education and Research Centre and the DBT-supported Bioinformatics Centre and Graphics facility. The work is supported by the DBT. MV is a Distinguished Biotechnologist Awardee of the DBT.

References

Begley, T. P., Kinsland, C. & Strauss, E. (2001). *Vitam. Horm.* **61**, 157–171.
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.
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.* D**54**, 905–921.
Cheek, S., Ginalski, K., Zhang, H. & Grishin, N. V. (2005). *BMC Struct. Biol.* **5**, 6–24.
Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Das, S., Kumar, P., Bhor, V., Surolia, A. & Vijayan, M. (2005). *Acta Cryst.* F**61**, 65–67.
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.
Daugherty, M., Polanuyer, B., Farrell, M., Scholle, M., Lykidis, A., de Crecy-Lagard, V. & Osterman, A. (2002). *J. Biol. Chem.* **277**, 21431–21439.
Engel, C. & Wierenga, R. (1996). *Curr. Opin. Struct. Biol.* **6**, 790–797.
Esnouf, R. (1997). *J. Mol. Graph.* **15**, 132–134.
Felsenstein, J. (1989). *Cladistics*, **5**, 164–166.
French, G. S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–525.
Genschel, U. (2004). *Mol. Biol. Evol.* **21**, 124–151.
Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). *Bioinformatics*, **19**, 163–164.
Hubbard, S. J. & Thornton, J. M. (1993). *NACCESS Computer Program.* Department of Biochemistry and Molecular Biology, University College, London.
Ivey, R. A., Zhang, Y. M., Virga, K. G., Hevener, K., Lee, R. E., Rock, C. O., Jackowski, S. & Park, H. W. (2004). *J. Biol. Chem.* **279**, 35622–35629.
Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
Kleywegt, G. J. & Jones, T. A. (1998). *Acta Cryst.* D**54**, 1119–1131.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Kupke, T., Hernandez-Acosta, P. & Culianez-Macia, F. A. (2003). *J. Biol. Chem.* **278**, 38229–38237.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leipe, D. D., Koonin, E. V. & Aravind, L. (2003). *J. Mol. Biol.* **333**, 781–815.
Leonardi, R., Zhang, Y. M., Rock, C. O. & Jackowski, S. (2005). *Prog. Lipid Res.* **44**, 125–153.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Navaza, J. (1994). *Acta Cryst.* A**50**, 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.
Saikrishnan, K., Kalapala, S. K., Varshney, U. & Vijayan, M. (2005). *J. Mol. Biol.* **345**, 29–38.
Saikrishnan, K., Manjunath, G. P., Singh, P., Jeyakanthan, J., Dauter, Z., Sekar, K., Muniyappa, K. & Vijayan, M. (2005). *Acta Cryst.* D**61**, 1140–1148.
Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.
Schneider, T. R. (2002). *Acta Cryst.* D**58**, 195–208.
Schüttelkopf, A. W. & van Aalten, D. M. (2004). *Acta Cryst.* D**60**, 1355–1363.
Song, W. J. & Jackowski, S. (1994). *J. Biol. Chem.* **269**, 27051–27058.
Terwilliger, T. C. (2000). *Nature Struct. Biol.* **7**, *Suppl.*, 935–939.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Vallari, D. S., Jackowski, S. & Rock, C. O. (1987). *J. Biol. Chem.* **262**, 2468–2471.
Vijayan, M. (1980). *Acta Cryst.* A**36**, 295–298.
Vijayan, M. (2005). *Tuberculosis*, **85**, 357–366.
Walshaw, J. & Woolfson, D. N. (2001). *J. Mol. Biol.* **307**, 1427–1450.
Yun, M., Park, C. G., Kim, J. Y., Rock, C. O., Jackowski, S. & Park, H. W. (2000). *J. Biol. Chem.* **275**, 28093–28099.
Zhang, Y. M., Rock, C. O. & Jackowski, S. (2005). *J. Biol. Chem.* **280**, 32594–32601.
Zhang, Y. M., Rock, C. O. & Jackowski, S. (2006). *J. Biol. Chem.* **281**, 107–114.
Zhou, B., Westaway, S. K., Levinson, B., Johnson, M. A., Gitschier, J. & Hayflick, S. J. (2001). *Nature Genet.* **28**, 345–349.