Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Characterization of the proline-utilization pathway in *Mycobacterium tuberculosis* through structural and functional studies

Thomas Lagautriere,[a] Ghader Bashiri,[a] Neil G. Paterson,[a]‡ Michael Berney,[b]§ Gregory M. Cook[b] and Edward N. Baker[a]*

[a]Structural Biology Laboratory, School of Biological Sciences and Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Auckland 1010, New Zealand, and [b]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand

‡ Current address: Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot OX11 0DE, England.
§ Current address: Department of Microbiology and Immunology, Albert Einstein College of Medicine, New York, NY 10467, USA.

Correspondence e-mail: ted.baker@auckland.ac.nz

Received 25 November 2013
Accepted 20 December 2013

PDB references: PruA (long cell), 4ids; PruA (short cell), 4idm; PruA (long cell), 4ihi; PruA (short cell), 4jdc

The proline-utilization pathway in *Mycobacterium tuberculosis* (*Mtb*) has recently been identified as an important factor in *Mtb* persistence *in vivo*, suggesting that this pathway could be a valuable therapeutic target against tuberculosis (TB). In *Mtb*, two distinct enzymes perform the conversion of proline into glutamate: the first step is the oxidation of proline into Δ^1-pyrroline-5-carboxylic acid (P5C) by the flavoenzyme proline dehydrogenase (PruB), and the second reaction involves converting the tautomeric form of P5C (glutamate-γ-semialdehyde) into glutamate using the NAD^+-dependent Δ^1-pyrroline-5-carboxylic dehydrogenase (PruA). Here, the three-dimensional structures of *Mtb*-PruA, determined by X-ray crystallography, in the apo state and in complex with NAD^+ are described at 2.5 and 2.1 Å resolution, respectively. The structure reveals a conserved NAD^+-binding mode, common to other related enzymes. Species-specific conformational differences in the active site, however, linked to changes in the dimer interface, suggest possibilities for selective inhibition of *Mtb*-PruA despite its reasonably high sequence identity to other PruA enzymes. Using recombinant PruA and PruB, the proline-utilization pathway in *Mtb* has also been reconstituted *in vitro*. Functional validation using a novel NMR approach has demonstrated that the PruA and PruB enzymes are together sufficient to convert proline to glutamate, the first such demonstration for monofunctional proline-utilization enzymes.

1. Introduction

Mycobacterium tuberculosis (*Mtb*) is the causative agent of tuberculosis (TB). Worldwide, TB is responsible for almost two million deaths per year, making *Mtb* one of the most devastating human pathogens (World Health Organization, 2011). Although reasonably effective drugs exist for TB treatment, they require prolonged multi-drug treatment regimens, leading to patient noncompliance. In addition, they are not effective against the nonreplicating persistent form of *Mtb*. The emergence of multidrug-resistant (MDR), extremely drug-resistant (XDR) and totally drug-resistant (TDR) *Mtb* strains, together with complications from co-infection with HIV/AIDS, requires the development of new and more effective anti-TB agents (Wayne & Sohaskey, 2001; Shah *et al.*, 2007). A detailed understanding of the biochemical and physiological state of the active and persistent *Mtb* is therefore of high priority in order to identify new therapeutic targets.

Proline metabolism has recently been implicated as an important factor in the adaptation of mycobacteria to slow growth rate and hypoxia (Berney *et al.*, 2012). It is believed that the proline-utilization pathway protects mycobacterial cells by detoxifying methylglyoxal, a toxic compound that can

Table 1
Primers used in the amplification of Rv1187 and Rv1188 constructs.

Construct	Primer sequences (5′–3′)	Restriction enzyme
Rv1187	Forward GGGATGCTGGCA**CATATG**GACGCGATCACCCAGGTGCCG	*Nde*I
	Reverse TGACAGCAACCG**AAGCTT**TCAGCGCTCGGCGCACCCCC	*Hin*dIII
Rv1188	Forward TTTCAGGGCG**CCATGG**CCGGCTGGTTCGCGCACACG	*Nco*I
	Reverse AGCAACCGGCG**AAGCTT**GCGCTCGGCGCACC	*Hin*dIII

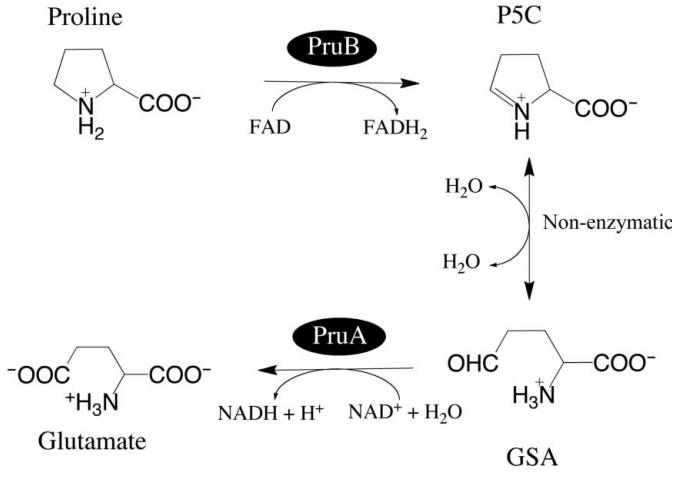


Figure 1
Proline utilization in *Mtb*. PruB irreversibly converts proline into Δ^1-pyrroline-5-carboxylic acid (P5C), which is in equilibrium with glutamate-γ-semialdehyde (GSA). Glutamate is then produced from GSA using PruA.

damage DNA and proteins within cells (Huang *et al.*, 2008; Kalapos, 1999). Methylglyoxal is produced as a by-product of endogenous glycerol metabolism (Berney *et al.*, 2012), but can react with Δ^1-pyrroline-5-carboxylic acid (P5C), an intermediate in the proline-utilization pathway, to form nontoxic 2-acetyl-1-pyrroline. Interestingly, the open reading frames (ORFs) encoding the enzymes involved in proline-utilization pathway are essential for *Mtb* growth (Berney & Cook, 2010; Berney *et al.*, 2012; Griffin *et al.*, 2011; Zhang *et al.*, 2012), implying that they possess important metabolic and physiological roles.

Proline is the only imino acid used in protein structures, and the incorporation of its α N atom within a pyrroline ring therefore demands a specific family of enzymes for its metabolism (Adams & Frank, 1980). The enzymes that mediate proline utilization differ between organisms, although the catalytic and mechanistic steps appear to remain similar (Phang *et al.*, 2008). Whereas some bacteria possess a bifunctional enzyme containing two catalytic domains that perform the conversion of proline into glutamate (Tanner, 2008; Lee *et al.*, 2003), in others the complete array of proline-utilization elements is incorporated into a trifunctional enzyme, in which an additional DNA-binding domain acts as a transcriptional regulator (*e.g. Escherichia coli* PutA). All eukaryotes and some bacteria, however, are restricted to monofunctional enzymes, with each enzyme having a unique catalytic function (Huang *et al.*, 2007; Inagaki *et al.*, 2006). In each case, the first step involves the conversion of proline into Δ^1-pyrroline-5-carboxylic acid (P5C), catalysed by proline dehydrogenase (*Mtb*-PruB), using FAD as coenzyme. P5C is in a tautomeric equilibrium with glutamate-γ-semialdehyde (GSA), which is subsequently converted into glutamate using the NAD^+-dependent Δ^1-pyrroline-5-carboxylate dehydrogenase (P5CDH; *Mtb*-PruA; Fig. 1).

PruA is a member of the aldehyde dehydrogenase family, and crystal structures of homologues are available from human (PDB entry 3v9g; Srivastava *et al.*, 2012), mouse (PDB entry 4lh3; Srivastava *et al.*, 2012) and several bacterial species [*Bacillus licheniformis* (PDB entry 3rjl; New York Structural Genomics Research Consortium, unpublished work), *B. halodurans* (PDB code 3qan; New York Structural Genomics Research Consortium, unpublished work) and *Thermus thermophilus* (PDB code 2eiw; Inagaki *et al.*, 2006)]. There are also reported PruB structures from *T. thermophilus* (PDB entry 2g37; White *et al.*, 2007), *Deinococcus radiodurans* (PDB entry 4h6r; Luo *et al.*, 2012) and *Pyrococcus horikoshii* (PDB entry 1y56; Tsuge *et al.*, 2005). No structure–function studies have been reported on the proline-utilization enzymes from *Mtb*, although a paper that was published during the preparation of our manuscript has described functional validation of the *Mtb*-PruB enzyme (Serrano & Blanchard, 2013).

Here, we present crystal structures of *Mtb*-PruA in the apo form and in complex with NAD^+ refined to 2.5 and 2.1 Å resolution, respectively. These structures show unexpected differences in the active site compared with other PruA enzymes, which suggest the possibility of selective inhibition of *Mtb*-PruA as a step towards future drug-design efforts against TB. In addition, we describe an *in vitro* reconstitution of the *Mtb* proline-utilization pathway using purified PruA and PruB enzymes, and the use of a novel NMR approach to verify that PruA and PruB work in cohort in *Mtb* to convert proline into glutamate.

2. Materials and methods

2.1. PCR amplification and cloning

The ORFs encoding PruA (Rv1187) and PruB (Rv1188) were amplified from *M. tuberculosis* H37Rv genomic DNA using *Pfu*Ultra II Fusion HS DNA Polymerase (Stratagene) and the primers outlined in Table 1. The amplified products were cloned into the pYUB28b vector (Bashiri *et al.*, 2010) using *Nde*I/*Hin*dIII restriction enzymes (Invitrogen) for PruA and *Nco*I/*Hin*dIII for PruB. The stop codon at the end of the PruB open reading frame was deleted to allow in-frame translation of a His_6 tag at the C-terminus. The constructs were separately transformed into *E. coli* Top10 electrocompetent cells and plated onto low-salt LB agar medium supplemented with 50 μg ml^{-1} hygromycin B. Positive clones were verified using colony PCR, followed by restriction digestion and sequencing.

2.2. Expression in *M. smegmatis*

Electrocompetent *M. smegmatis* mc^{2}4517 (Wang *et al.*, 2010) cells (40 µl) were mixed with 2 µl of plasmid DNA and incubated on ice for 5 min. 260 µl glycerol [10%(*v*/*v*)] was added to the mixture, after which the cells were electroporated in a 0.2 cm cuvette (electroporation parameters: R = 1000 Ω, Q = 25 µF and V = 2.5 kV; BioRad Gene Pulser). The transformed cells were immediately resuspended in 1 ml 7H9/ADS/Tween 80 solution (Difco and BBL Middlebrook) and were grown for 3 h at 310 K. The cells were then plated onto LB agar medium supplemented with 50 µg ml^{-1} each of kanamycin and hygromycin B and incubated for 4 d at 310 K.

Protein expression was performed in autoinduction medium using previously published protocols (Studier, 2005). A single colony was picked and used to inoculate a starter culture in non-inducing MDG medium (25 m*M* Na_2HPO_4, 25 m*M* KH_2PO_4, 50 m*M* NH_4Cl, 5 m*M* Na_2SO_4, 2 m*M* $MgSO_4$, 0.5% D-glucose, 0.25% L-aspartate, 0.2× metal mix and 0.05% Tween 80). The starter culture was incubated at 310 K for 3 d and was then used to inoculate (1:100 dilution) an autoinduction culture in ZYM-5052 medium (1% tryptone, 0.5% yeast extract, 25 m*M* Na_2HPO_4, 25 m*M* KH_2PO_4, 50 m*M* NH_4Cl, 5 m*M* Na_2SO_4, 2 m*M* $MgSO_4$, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose, 1× metal mix and 0.05% Tween 80). The expression was carried out for 4 d at 310 K for an optimum expression level.

Selenomethionine-substituted PruA (SeMet-PruA) was also expressed in *M. smegmatis* using autoinduction medium following a published protocol for selenium incorporation in this host organism (Bashiri *et al.*, 2007). SeMet-PruA was expressed in the selenomethionine-supplemented minimal medium PASM-5052 [50 m*M* Na_2HPO_4, 50 m*M* KH_2PO_4, 25 m*M* $(NH_4)_2SO_4$, 2 m*M* $MgSO_4$, 1× trace metals, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose, 200 µg ml^{-1} of each of 17 amino acids (no cysteine, tyrosine or methionine), 10 µg ml^{-1} methionine, 125 µg ml^{-1} selenomethionine, 100 n*M* vitamin B_{12} and 0.05% Tween 80]. Expression was maintained for 14 d at 310 K.

2.3. Purification of native and SeMet-substituted *Mtb*-PruA

Mtb-PruA was cloned into the pYUB28b vector with an N-terminal His_6 tag, facilitating the subsequent purification steps. Both native and SeMet-PruA proteins were purified from *M. smegmatis* cells using the same procedure, as follows. The cells were harvested and resuspended in 20 m*M* Tris–HCl pH 8.5, 150 m*M* NaCl, 20 m*M* imidazole, 1 m*M* β-mercaptoethanol (β-ME) and then lysed by two passes through a cell disrupter (Microfluidizer M-110P). The lysate was centrifuged at 20 000*g* to separate the insoluble material. The recombinant protein was first purified using an immobilized metal-affinity chromatography (IMAC) step by loading the supernatant onto a pre-packed TALON resin column (Clontech) that had been pre-equilibrated in the lysis buffer. The column was washed with the lysis buffer until no protein was detected in the flowthrough using a Bio-Rad protein assay. The protein was subsequently eluted with buffer containing 250 m*M* imidazole and the purity was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE). The elution fraction was concentrated using a Vivaspin centrifugal concentrator with a 30 kDa molecular-weight cutoff, and then injected onto a size-exclusion Superdex 200 10/30 column (GE Healthcare) pre-equilibrated in 20 m*M* Tris–HCl pH 8.5, 150 m*M* NaCl, 1 m*M* β-ME. The His_6 tag was retained on the protein during subsequent crystallization trials and functional experiments.

2.4. Purification of *Mtb*-PruB

Mtb-PruB was cloned into the pYUB28b vector with a C-terminal His_6 tag. All harvesting and purification steps were carried out at 277 K. The cells expressing the *Mtb*-PruB protein were harvested and resuspended in 50 m*M* HEPES pH 8.0, 100 m*M* sodium formate, 20 m*M* imidazole, 0.6 m*M* tris(2-carboxyethyl)phosphine (TCEP), 1 m*M* *n*-octyl-β-D-glucoside (BOG). Cell lysis and centrifugation steps were performed as previously described for PruA, and the supernatant was then loaded onto a HisTrap FF 5 ml nickel-affinity column (GE Healthcare) pre-equilibrated with the lysis buffer. After a wash step, PruB was eluted using a gradient of imidazole in the buffer. The protein eluted at approximately 300 m*M* imidazole and was then injected onto a size-exclusion column pre-equilibrated in 50 m*M* HEPES pH 8.0, 100 m*M* sodium formate, 0.6 m*M* TCEP, 1 m*M* BOG.

2.5. *Mtb*-PruA crystallization and data collection

The native PruA protein was concentrated to 15 mg ml^{-1} (Vivaspin, 30 kDa molecular-weight cutoff) and crystals were grown using the sitting-drop vapour-diffusion technique at 291 K. Initial crystallization hits were obtained using a Cartesian nanolitre dispensing robot (Genome Solutions) and in-house crystallization screens (Moreland *et al.*, 2005). The successful crystallization condition was obtained and reproduced from the Morpheus protein crystallization screen (Molecular Dimensions; Gorrec, 2009), with the precipitant solution comprising 0.1 *M* Bicine–Tris pH 8.1, 12% polyethylene glycol (PEG) 1000, 12% PEG 3350, 12% 2-methyl-2,4-pentanediol (MPD) and 0.03 *M* each of sodium nitrate, disodium hydrogen phosphate and ammonium sulfate. The crystallization drops were formed by mixing 2 µl protein solution (PruA in 20 m*M* Tris–HCl pH 8.5, 150 m*M* NaCl, 1 m*M* β-ME) with 2 µl reservoir solution. SeMet-PruA crystals were grown using the same conditions. In addition, in order to obtain the structure of PruA bound to its potential cofactor, the native protein was co-crystallized with 250 µ*M* NAD^+ using the same crystallization conditions.

Considering that all PruA crystals were formed using the Morpheus crystallization screen (Gorrec, 2009), no additional cryoprotection step was required and the crystals were cooled directly in liquid nitrogen. X-ray diffraction data sets were collected from single crystals on the Australian Synchrotron MX2 beamline using an ADSC Quantum 315r CCD detector. Images were collected by rotating the crystal through 360° with 1° oscillations, each with a 1 s exposure time. Data-

Table 2
Data-collection statistics.

Values in parentheses are for the outermost resolution shell.

	SeMet-PruA (short)					
Data set	Remote	Inflection	Peak	Apo PruA (long)	PruA–NAD$^+$ (long)	PruA–NAD$^+$ (short)
Wavelength (Å)	0.953698	0.979231	0.979083	1.5418	0.953692	0.953700
Space group	*P*622	*P*622	*P*622	*P*622	*P*622	*P*622
Unit-cell parameters						
a (Å)	162.9	163.0	163.1	164.4	163.9	163.1
b (Å)	162.9	163.0	163.1	164.4	163.9	163.1
c (Å)	96.2	96.3	96.4	194.0	194.0	96.2
$\alpha = \beta$ (°)	90	90	90	90	90	90
γ (°)	120	120	120	120	120	120
Resolution (Å)	2.50 (2.60–2.50)	2.50 (2.60–2.50)	2.50 (2.60–2.50)	2.00 (2.15–2.00)	2.25 (2.37–2.25)	1.60 (1.69–1.60)
$R_{\rm merge}$	0.169 (0.542)	0.200 (0.725)	0.189 (0.734)	0.240 (1.664)	0.184 (0.570)	0.103 (1.981)
$R_{\rm p.i.m.}$	0.073 (0.235)	0.086 (0.318)	0.080 (0.320)	0.046 (0.338)	0.053 (0.170)	0.024 (0.548)
Unique reflections	24660	24323	24376	99133	73091	98328
Observed reflections	280615	279178	280960	2777368	1004614	1766636
Mean $I/\sigma(I)$	11.2 (4.3)	10.1 (3.1)	10.9 (3.1)	16.5 (2.6)	10.1	17.7
Multiplicity	11.4	11.5	11.5	28.0	13.7	18.0
Completeness (%)	94.2	93.2	93.1	99.7	99.9	99.8
CC_I mean†	0.903	0.802	0.802	0.702	0.958	0.581
MapCC org/inv‡ (%)	40.4/28.3	40.4/28.3	40.4/28.3	—	—	—

† Pearson correlation coefficient. ‡ Map correlation coefficients for original and inverted heavy-atom configurations.

Table 3
Structure-refinement statistics.

	SeMet-PruA (short)	Apo PruA (long)	PruA–NAD$^+$ (long)	PruA–NAD$^+$ (short)
PDB code	4idm	4ids	4ihi	4jdc
Resolution range (Å)	39.78–2.50	19.95–2.04	29.51–2.25	141.29–1.60
No. of reflections	39636	98996	68453	93414
$R_{\rm work}/R_{\rm free}$ (%)	14.4/20.7	22.9/24.1	27.7/30.7	17.1/20.2
No. of atoms				
Protein	4110	4142	4109	4112
Ligand	—	—	44	27
Water	223	490	450	389
R.m.s.d. from ideality				
Bonds (Å)	0.017	0.01	0.02	0.03
Angles (°)	1.80	1.03	1.07	2.64
Average B factors (Å^2)				
Protein	26.13	26.74	16.67	24.77
Waters	29.68	39.21	28.30	36.81
Ramachandran statistics				
Favoured (%)	96.23	97.76	96.80	97.18
Allowed (%)	3.77	2.24	3.01	2.63
Outliers (%)	0	0	0.19	0.19
Poor rotamers (%)	2.64	1.20	1.45	1.68
MolProbity score; percentile	1.62; 99th	1.09; 100th	1.55; 98th	1.61; 82nd

collection statistics and unit-cell data are summarized in Table 2.

2.6. *Mtb*-PruA structure determination and refinement

The native and SeMet-PruA data sets were indexed and processed using *XDS* (Kabsch, 2010), re-indexed using *POINTLESS* and scaled with *SCALA* from the *CCP*4 program suite (Winn *et al.*, 2011; Evans, 2006). SeMet-PruA data sets collected at three different energies were combined as a single file using *POINTLESS* and scaled using *SCALA* with a maximum resolution of 2.5 Å. The scaled unmerged data were input into the *SHELX* suite of programs for structure solution. Analysis of the distribution of anomalous signal [$I/\sigma(I)$ *versus* resolution] using *SHELXC* showed the presence of significant anomalous signal to 2.5 Å resolution, which was then chosen as the resolution cutoff for structure determination. All eight SeMet atoms were found using *SHELXD* (Schneider & Sheldrick, 2002), with occupancies varying between 1 and 0.63, giving an overall correlation coefficient of 64%. *SHELXE* was then used for phasing and density modification, in which the output showed a clear difference between the original and inverted solutions (Sheldrick, 2010; Morris *et al.*, 2003). The resulting electron-density map was directly suitable for input into the *ARP*/*wARP* automated model-building protocols (Morris *et al.*, 2003). The structures of native apo PruA and PruA–NAD$^+$ were then solved by molecular replacement using *MOLREP* (Vagin & Teplyakov, 2010) with the SeMet-PruA structure as a search model. All structures were refined by cycles of manual building using *Coot* (Emsley & Cowtan, 2004) and refinement using *BUSTER* (Bricogne, 1993). Full refinement statistics are shown in Table 3.

2.7. Activity assays

For UV–visible spectrophotometry experiments, the PruA and PruB enzymes were dialyzed against assay buffer (20 m*M* potassium phosphate buffer pH 8.0, 100 m*M* NaCl, 0.2 m*M* TCEP). NAD$^+$ stock solution was prepared in the assay buffer and used at a final concentration of 0.5 m*M*. The proteins and the cofactor were mixed together in the assay buffer and the reaction was started by the addition of different concentrations of proline over the range 0.2 µ*M*–1 m*M*. Different PruA:PruB ratios were used to find the optimal ratio for the assays. The reactions were monitored by measuring the increase in absorbance at 340 nm using a Cary 4000 UV–visible spectrophotometer (Varian) at ambient temperature.

The increase in absorbance measures the conversion of NAD^+ to NADH in the PruA-catalysed reaction.

For NMR spectroscopy experiments, both the proteins and the cofactor were prepared as described previously for the UV–Vis spectrophotometry experiments. ^{15}N-labelled proline and glutamate were purchased from Cambridge Isotope Laboratories. An optimal PruA:PruB ratio of 1:10 was used for the assays, corresponding to concentrations of 0.3 µ*M* PruA and 3 µ*M* PruB. The reactions were set up as above and were supplemented with 10% D_2O, and were then started by the addition of ^{15}N-labelled proline to a final concentration of 1 m*M*. Activity was monitored using two-dimensional ^{1}H–^{15}N NMR spectroscopy at 291 K for 18 h. The relative ratios of the substrate, the intermediate and the product were determined by dividing the integrals of the corresponding signals over the sum of the integrals at any given time point. NMR spectra were collected on a Bruker DRX 600 MHz instrument equipped with a 5 mm *z*-gradient TCI cryoprobe. Long-range correlations between coupled ^{1}Hβ protons and ^{15}N were obtained from the analysis of ^{15}N–^{1}H HMBC data. Data were processed and analysed using *NMRPipe* (Delaglio *et al.*, 1995) and *CcpNmr Analysis* (Vranken *et al.*, 2005).

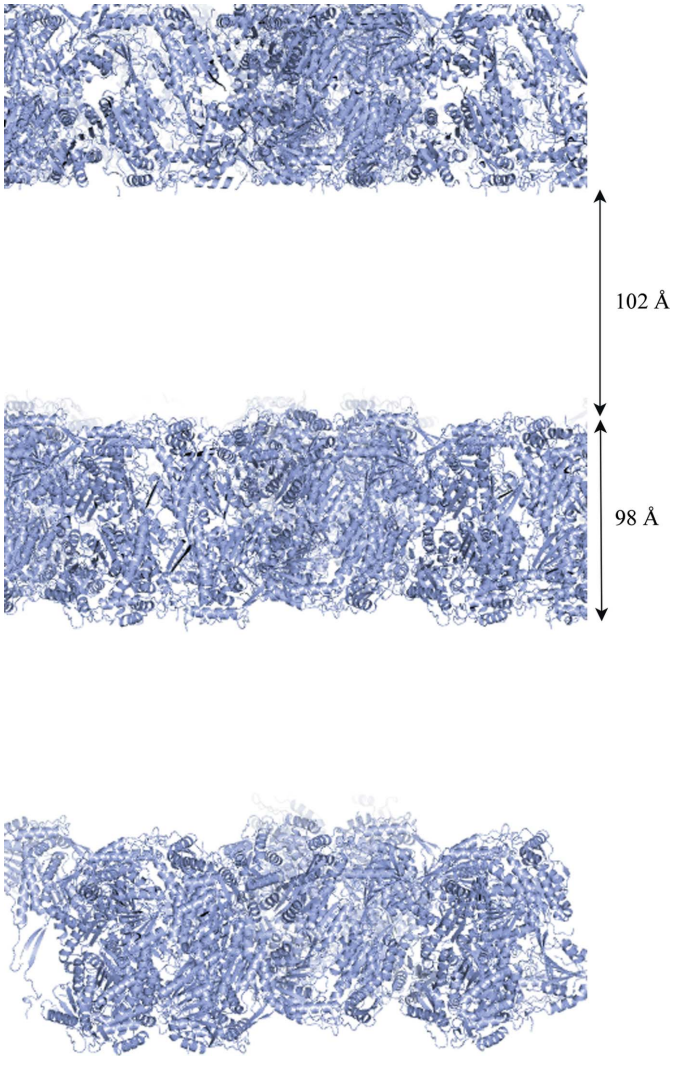


Figure 2
PruA crystal packing in the long unit-cell dimension along the *c* axis. The overall crystal packing is shown with an arrangement of layers of ordered molecules separated by empty layers of 102 Å. An ordered layer of protein molecules could be placed within the empty layers.

3. Results and discussion

3.1. *Mtb*-PruA expression, purification and crystallization

Mtb-PruA was expressed as a soluble protein in *M. smegmatis* using autoinduction medium. Both native and SeMet-PruA proteins were purified using IMAC and size-exclusion chromatography steps, and hexagonal crystals were grown using the Morpheus crystallization screen (Gorrec, 2009) within 3 d at 291 K.

3.2. Structure determination

During the crystallization of *Mtb*-PruA, two different crystal forms with the same space group (*P*622) were obtained within all crystallization drops. These crystal forms differed in that the unit-cell dimension along the *c* axis was almost doubled in the longer unit cell (194 *versus* 96 Å; Table 2). An initial structure for PruA was obtained using multiwavelength anomalous dispersion (MAD) phasing from a short-cell SeMet-PruA crystal with one PruA molecule per asymmetric unit. The substructure solution and phase calculations were performed using the *SHELX* suite, after which the model was built using the automated model-building program *ARP/wARP*, which built 98% of the model without any manual refinement. This SeMet-PruA structure was refined using data to 2.5 Å resolution, resulting in a crystallographic *R* factor of 14.4% and an $R_{\rm free}$ of 20.7% (Table 3). The structure of a long-cell crystal of native apo PruA was then solved by molecular replacement using the small-cell SeMet-PruA structure as the search model, and was refined using 2.0 Å resolution data to a crystallographic *R* factor of 22.9% and an $R_{\rm free}$ of 24.1% (Table 3).

PruA–NAD^+ crystals were also obtained in both short-cell and long-cell crystal forms using the co-crystallization protocol described earlier, and their structures were determined by molecular replacement. The short-cell PruA–NAD^+ structure was refined at 1.6 Å resolution, with a final crystallographic *R* factor of 17.1% and an $R_{\rm free}$ of 20.2%, whereas the structure with the long unit cell was refined at 2.1 Å resolution, resulting in a crystallographic $R_{\rm factor}$ of 27.7% and an $R_{\rm free}$ of 30.7% (Table 3).

3.3. Crystal disorder

The two crystal forms obtained differ in that the long-cell crystals have an unusual arrangement of alternating crystal layers along the *c* axis (Fig. 2), in which an ordered layer is followed by a disordered layer such that layers with ordered molecules are separated by 102 Å gaps in which no ordered structure can be seen. The nature of this disorder is discussed later, but we note here that the electron density in the missing layers is weak and has an unusual distribution. Although it shows some features of connectivity and parts of secondary-

structural elements could be recognized, neither *Phaser* nor *MOLREP* was able to find any molecules within these layers. The width of the ordered layer is ∼98 Å, suggesting that a layer of ordered molecules could be accommodated in the 102 Å empty layer.

The *Mtb*-PruA apo structure determined from the short-cell crystals is essentially the same as that determined from the ordered part of the long-cell crystals, with a root-mean-square difference (r.m.s.d.) of only 0.38 Å over 537 C^{α} atomic positions. The PruA–NAD^+ complexes, however, show a surprising difference, whereby the crystals with the long cell show an ordered active site in which NAD^+ can be fully modelled, whereas the crystals with the shorter unit cell display a disordered active site in which only partial density is observed for NAD^+, with no apparent electron density for the nicotinamide-riboside moiety. This is discussed further below.

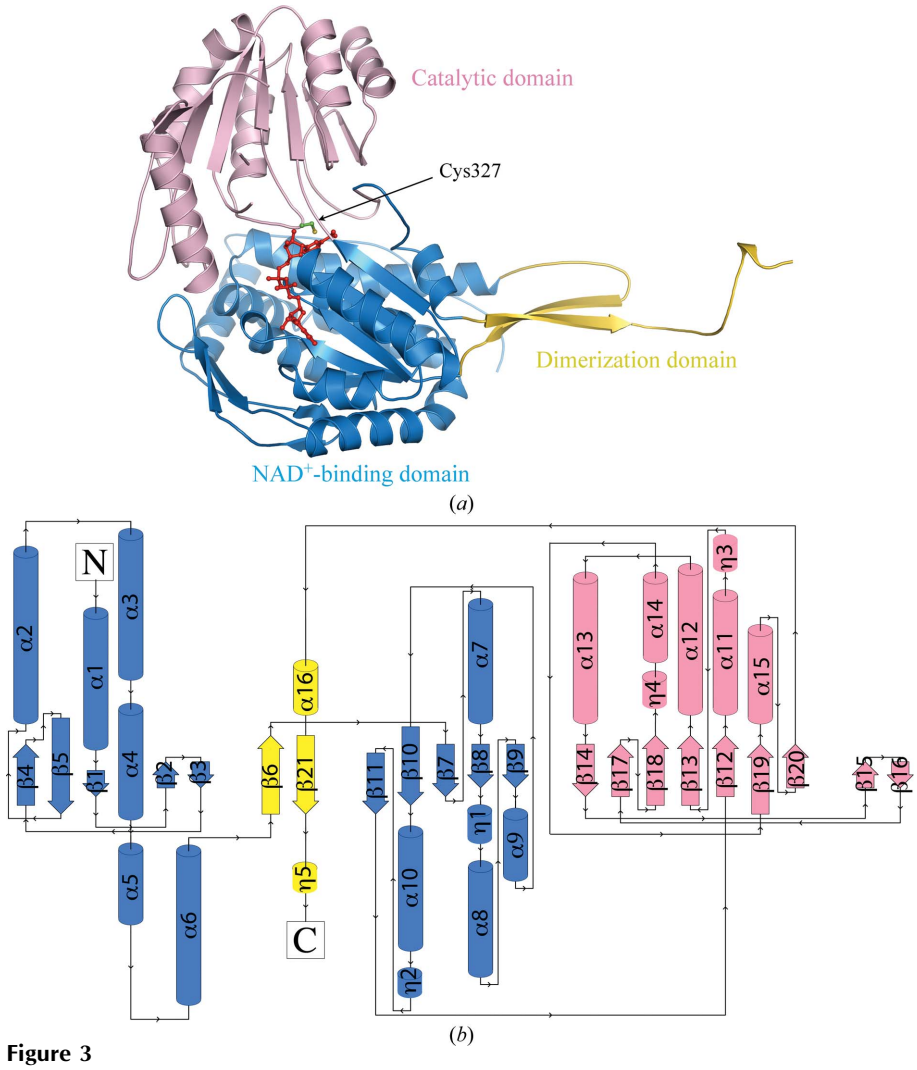


Figure 3
Three-dimensional crystal structure of *Mtb*-PruA. (*a*) A cartoon representation of the PruA monomer structure with NAD^+ shown as red sticks and the catalytic cysteine Cys327 shown as green sticks. The structure displays a catalytic domain (pink), an NAD^+-binding domain (blue) and a dimerization domain (yellow). (*b*) PruA topology diagram with the domains coloured the same as in (*a*).

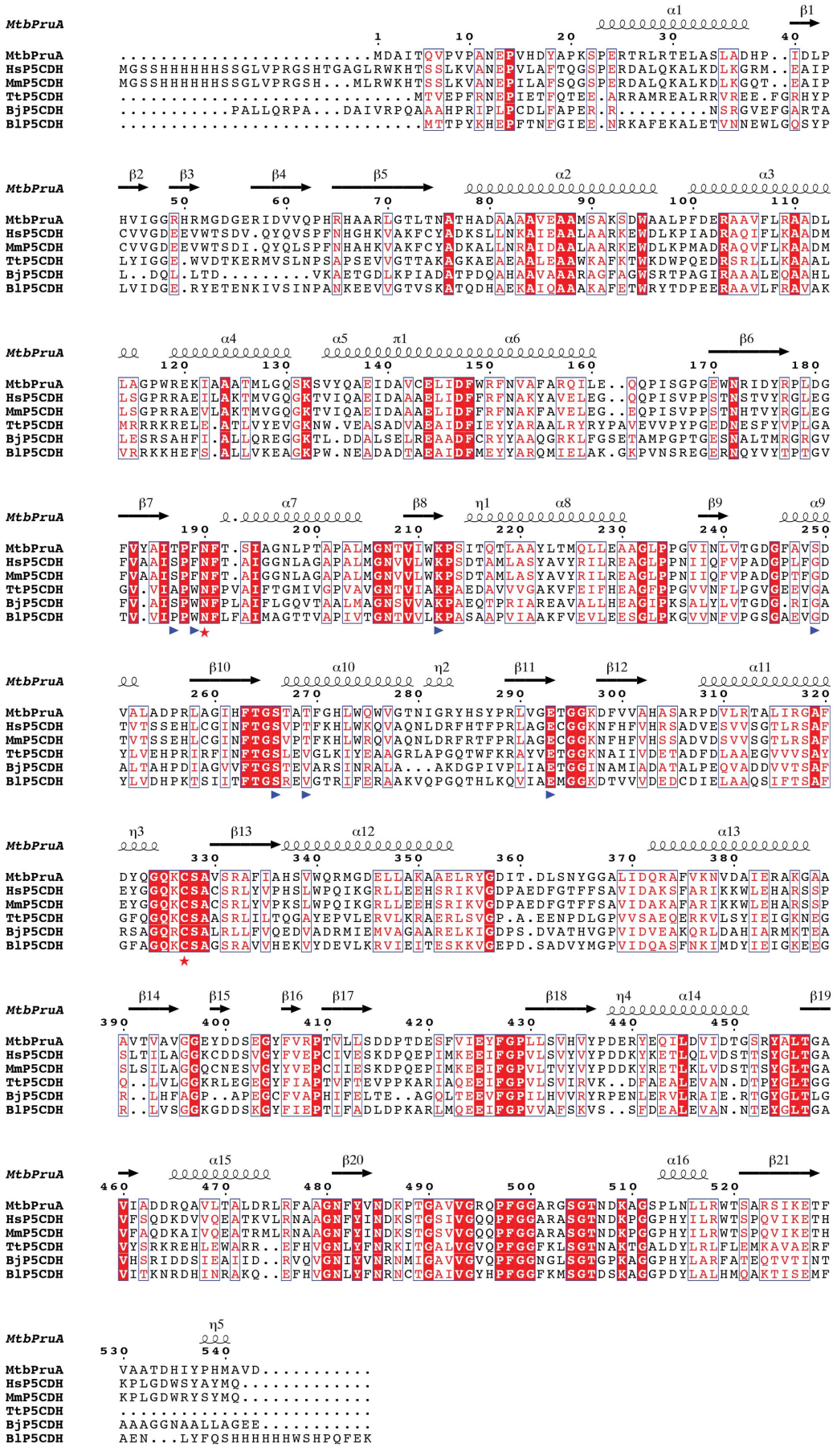
3.4. Overall structure of *Mtb*-PruA

The ordered short-cell crystal structure contains one molecule of *Mtb*-PruA in the asymmetric unit. The whole molecule, comprising residues 1–543, is well defined and has excellent electron density except for residues 419–425 and residue 490, for which no electron density was observed. These residues are located on the protein surface with no apparent crystal contacts. Residues 419–425, however, are ordered in the structure with the long unit cell. In each of the PruA structures the Ramachandran plot has a single outlier at Phe498, which is located after a proline residue and is well defined by the electron density. This residue is central to a key conformational difference in the *Mtb*-PruA active site when compared with other PruA enzymes, as described later.

The fold of *Mtb*-PruA is organized into three domains: an N-terminal nucleotide (NAD^+) binding domain (residues 1−160, 181–294 and 502−520), a catalytic domain (residues 295–501) and a dimerization domain (residues 161−180 and 521−543) (Fig. 3). Both the fold, and sequence comparisons (Fig. 4), clearly identify *Mtb*-PruA as a member of the very large aldehyde dehydrogenase (ALDH) family (Sophos & Vasiliou, 2003). The NAD^+-binding domain adopts a modified Rossmann fold that is characteristic of the ALDH family; this binds the NAD^+ cofactor at the C-terminal face of the β-sheet. The catalytic domain, which includes the catalytic cysteine (Cys327), also has a Rossmann-type α/β fold, but with seven β-strands. The active site is located at the interface between the NAD^+-binding and the catalytic domains, with a relatively large water-filled channel leading to the catalytic Cys327. The dimerization domain has two β-strands (β4 and β17) that project out from the monomer, together with a C-terminal extension of ∼20 residues, which is not seen in most ALDH family members.

3.5. Structural comparisons

A search of the Protein Data Bank (PDB) using *SSM* (Krissinel & Henrick, 2004) shows that the most similar homologues of *Mtb*-PruA are all monofunctional P5C dehydrogenase (P5CDH) enzymes from a variety of species. These include the very similar human and mouse enzymes (*Hs*-P5CDH and *Mm*-P5CDH; PDB entries 3v9g and 4lh3; Srivastava *et al.*, 2012), with 45% and 44% sequence identity, respectively, and r.m.s.d.s of ∼1.30 Å over 520 residues,



together with bacterial enzymes from *T. thermophilus* (*Tt*-P5CDH; PDB entry 2eiw; Inagaki *et al.*, 2006), *B. halodurans* (PDB entry 3qan) and *B. licheniformis* (PDB code 3rjl). The bacterial enzymes all have ∼30% sequence identity to *Mtb*-PruA and r.m.s.d.s of 1.94–1.98 Å over ∼500 residues. It is curious that the most similar enzymes to the *Mtb*-PruA are the two mammalian enzymes. These share not only the same monomer fold with *Mtb*-PruA but also the same dimer association formed through a similar C-terminal extension, as described below.

Searches of the PDB also match *Mtb*-PruA with the P5CDH domains from the multifunctional proline-utilization enzymes from three organisms: *Bradyrhizobium japonicum* (PDB entry 3haz; Srivastava *et al.*, 2010; 27% sequence identity and an r.m.s.d. of 1.9 Å over 445 aligned residues), *Geobacter sulfurreducens* (PDB entry 4f9i; 25% identity and an r.m.s.d. of 1.9 Å over 488 residues) and *Anabaena variabilis* (PDB entry 4h7n; 19% identity and an r.m.s.d. of 2.4 Å over 433 residues). The main structural difference between

Figure 4
Multiple sequence alignment of *Mtb*-PruA with P5CDH homologues. The catalytic residues Asn190 and Cys327 are indicated by red stars and residues involved in NAD^+ binding are indicated by blue triangles. MtbPruA, *Mycobacterium tuberculosis* PruA; HsP5CDH, human P5CDH; MmP5CDH, mouse P5CDH; TtP5CDH, *Thermus thermophilus* P5CDH; BjP5CDH, *Bradyrhizobium japonicum* P5CDH; BlP5CDH, *Bacillus licheniformis* P5CDH.

monofunctional and multifunctional P5CDH enzymes appears to be the absence of secondary-structure elements equivalent to strands $\beta1$–$\beta4$ from the N-terminal region of *Mtb*-PruA.

3.6. The PruA dimer

The crystal structure of *Mtb*-PruA shows that two PruA monomers assemble to form a highly stable dimer, in which 4146 Å^2 or 18% of the solvent-accessible surface area of each monomer is buried, as analysed by the program *PISA* (Krissinel & Henrick, 2007). This large dimer interface is consistent with the results from analytical size-exclusion chromatography and dynamic light scattering, which show that *Mtb*-PruA is dimeric in solution (data not shown). The dimer is substantially stabilized by arm swapping (Fig. 5), in which two β-strands of the dimerization domain from one monomer ($\beta6$ and $\beta21$) associate with the catalytic domain of the other monomer, resulting in the formation of an intermolecular β-sheet comprising nine β-strands. The C-terminal extension of the dimerization domain also interacts with the outer surface of the catalytic domain from the other monomer, further stabilizing the dimer. Although the C-terminal extension seen in *Mtb*-PruA is specific to only a few ALDH members, the arm swapping is a feature of dimer formation by all ALDH members.

The *Mtb*-PruA dimer is essentially the same as that formed by other P5CDH enzymes. Superposition of the *Mtb*-PruA dimer onto that formed by the human enzyme shows an r.m.s.d. in C$^\alpha$ atomic positions over 1048 residues of the dimer of 1.52 Å, which is only slightly greater than the r.m.s.d. when the monomers are compared (1.32 Å). There are, however, differences in detail. In particular, the buried hydrophobic surface between monomers is enhanced in *Mtb*-PruA owing to a local conformational difference involving residues 497–504 that appears to be specific to the *Mtb* enzyme. In the latter, Phe498 is buried in the dimer interface, surrounded by Leu514 from the same monomer and Ile174, Tyr176, Leu517 and Ala522 from the opposing monomer. In the other P5CDH enzymes, however, the equivalent Phe residue is instead located internally in the active site, and there are a number of accompanying changes in this part of the dimer interface that make it distinctly more polar; for example, Leu514, which contacts Phe498 in the *Mtb* enzyme, changes to Asp in the *T. thermophilus* enzyme and His in the mouse enzyme. The consequences for the active site are described below (§3.8).

3.7. NAD$^+$ binding

NAD$^+$ binds in the active site in an extended conformation (Fig. 6*a*), binding to the C-terminal face of the β-sheet formed by strands $\beta7$–$\beta11$. Superposition of the apo PruA and PruA–NAD$^+$ structures from either the long-cell or short-cell crystal forms shows that no large-scale conformational changes accompany NAD$^+$ binding, with the only significant movement being in the loop carrying Cys327, which moves to accommodate the nicotinamide ring of NAD$^+$. The overall pairwise r.m.s.d. over 537 C$^\alpha$ atom positions for the four apo PruA and PruA–NAD$^+$ structures is 0.27–0.40 Å.

The adenine ring of NAD$^+$ occupies a mostly hydrophobic pocket formed by Tyr184, Ile186, Phe245, Thr269, His272 and Leu273, all from the nucleotide-binding domain, and makes just one direct hydrogen bond, from N1 to the Ser249 hydroxyl group. The adenyl ribose group binds through its 2′-OH and 3′-OH groups to the amino group of Lys212, and through its 2′-OH group to the carbonyl O atom of Thr187. The central diphosphate moiety of the nucleotide binds in the same spatial location as in other Rossmann-fold enzymes, adjacent to the $\beta7$–$\alpha7$ loop, but this nonclassical Rossmann domain lacks the typical G*X*G*XX*G motif, and only a single hydrogen bond is made *via* this loop: from the peptide NH of Phe189 to one of the β-phosphate O atoms. Two other direct hydrogen bonds are present between an α-phosphate O atom and Ser266 (OG and NH) from the $\beta10$–$\alpha10$ loop. The nicotinamide ribose makes just a single direct hydrogen bond to the protein, from its 2′-OH to the carbonyl O atom of the catalytic Cys327, but is in van der Waals contact with Gly295 and Phe427. Finally, the key nicotinamide ring is sandwiched between the side chains of Thr264 and the catalytic Cys327. This requires movement of Cys327 to allow sufficient space for the nicotinamide, a phenomenon also observed when NAD$^+$ is bound by *Mm*-P5CDH and *Tt*-P5CDH. This adjustment places the sulfhydryl group of Cys327 only 3.1 Å from C4, the site of hydride transfer on the nicotinamide ring.

The short-cell and long-cell *Mtb*-PruA–NAD$^+$ complexes show an intriguing difference that is related to the need for Cys327 to move to accommodate the nicotinamide ring. In the long-cell structure the NAD$^+$ molecule is fully ordered and can be visualized completely as described above. In the short-cell structure, however, Cys327 is modified by β-ME, with the formation of a hydroxyethyl disulfide moiety (Cme327; Figs. 6*b* and 6*c*). This evidently cannot be displaced and no density can be seen for the nicotinamide ring and ribose, although the adenosyl diphosphate portion is bound identically.

Overall, NAD$^+$ binding in *Mtb*-PruA is characterized by a rather small number of direct hydrogen bonds, nine in total, plus another ten water-mediated interactions. This represents a higher proportion (53%) of water-mediated interactions

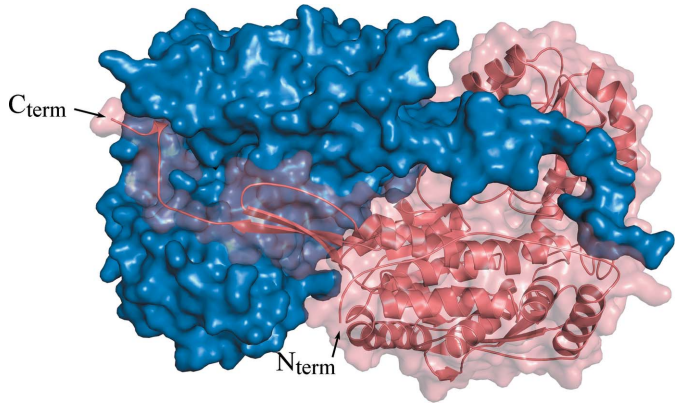


Figure 5
The *Mtb*-PruA dimer. Two monomers are associated through arm swapping to form the dimer. Both monomers are displayed as surface representations, but with one monomer also shown as a cartoon model.

than is usual for NAD$^+$-binding sites within Rossmann-fold enzymes, which is typically about 30% (Bottoms *et al.*, 2002). The relative lack of direct hydrogen bonds is consistent with the role of NAD$^+$ as a co-substrate that dissociates from the enzyme during its catalytic reaction.

3.8. Catalysis, substrate binding and prospects for inhibitor development

The PruA catalytic reaction can be understood in terms of a series of steps that begin with binding of the coenzyme NAD$^+$ and substrate GSA. Nucleophilic attack on the C$^\delta$ atom of GSA by the active-site Cys327, with the formation of a thiohemiacetal intermediate, is followed by hydride transfer to C4 of the nicotinamide ring and dissociation of the resulting NADH. Finally, hydrolysis of the thioacyl intermediate leads to release of the product glutamate.

The structural elements of PruA catalysis have mostly been elucidated through structural analyses of NAD$^+$ and glutamate complexes of *Tt*-P5CDH (Inagaki *et al.*, 2006) and *Mm*-P5CDH (Srivastava *et al.*, 2012). *Mtb*-PruA shares most of the same structural elements. Superposition of the NAD$^+$ complexes of *Mtb*-PruA, *Tt*-P5CDH and *Mm*-P5CDH shows that NAD$^+$ binds identically in each case. The main-chain NH of Cys327 and the side chain of Asn190 are in position to provide the oxyanion hole that binds the carbonyl O atom of GSA and stabilizes the developing negative charge. The α-carboxyl and α-amino groups of GSA should be bound by the main-chain NH groups of Gly477 and Ala488 and the hydroxyl of Ser328 by analogy with the glutamate binding mode to *Tt*-P5CDH and *Mm*-P5CDH (Fig. 7).

Despite the conservation of active-site residues, there do appear to be structural differences that could be exploited for *Mtb*-specific inhibitor design. The most significant of these involves the binding site for the substrate, GSA. Although GSA is itself unstable, complexes of *Mm*-P5CDH and *Tt*-P5CDH with the product glutamate clearly define the probable substrate-binding site (Srivastava *et al.*, 2012; Inagaki *et al.*, 2006). This involves two Phe residues, Phe212 and Phe520 in *Mm*-P5CDH, equivalent to Phe191 and Phe498 in *Mtb*-PruA. In the mouse and *T. thermophilus* enzymes the two Phe side chains provide a pre-formed pocket, being in the same position in the apo form and in the NAD$^+$ and glutamate complexes. The glutamate is sandwiched between the two aromatic rings, while making hydrogen bonds to other residues through its α-amino and α-carboxyl groups. Despite the fact that all of the glutamate-binding residues are fully conserved at the sequence level, in *Mtb*-PruA Phe498 is flipped ~11.5 Å out of the active site into the dimer interface, as described earlier (§3.6). This leaves a substantial water-filled cavity in the active site, much larger than in the other P5CDH enzymes, which could be exploited for the development of *Mtb*-specific inhibitors.

The conformational difference that removes Phe498 from the active site is a consistent feature of all of the *Mtb*-PruA structures and was well tested during their structure determination and refinement. The short-cell apo structure was built *de novo* using experimental (MAD) phases, which gave very clear density for this part of the structure. We also removed residues 497–504 from each of the four structures at a

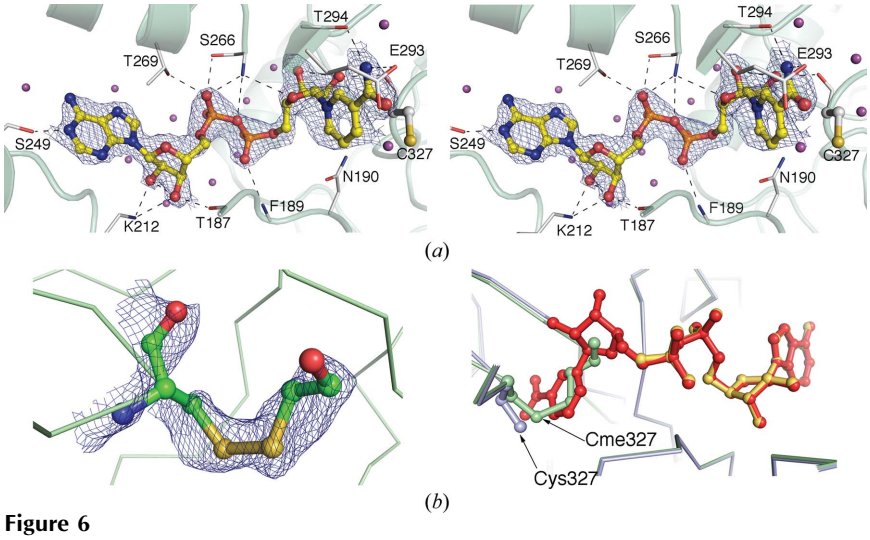


Figure 6
NAD$^+$ binding to *Mtb*-PruA. (*a*) Stereoview of NAD$^+$ binding in the ordered PruA active site of the long-cell crystal form. NAD$^+$ is shown in yellow, fitted to its $2F_o - F_c$ OMIT electron density contoured at 1.5σ. Water molecules are shown as purple spheres. (*b*) Cys327 in the short-cell crystal form, chemically modified as a hydroxyethyldisulfide derivative. The corresponding $2F_o - F_c$ OMIT electron density is also shown contoured at 1.5σ. (*c*) Superimposition of the NAD$^+$ molecules from the ordered and disordered *Mtb*-PruA active sites. In the long-cell structure, with unmodified Cys327, the NAD$^+$ molecule could be fully modelled (red sticks). In the short-cell structure, where Cys327 is modelled as a hydroxyethyldisulfide derivative (green sticks), NAD$^+$ could only be partially modelled (yellow sticks), as no electron density is apparent for the nicotinamide-riboside moiety.

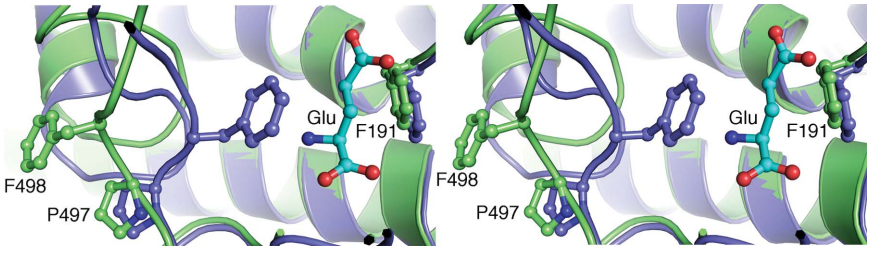


Figure 7
Conformational differences in the *Mtb*-PruA active site. Stereoview comparing the glutamate-binding site of mouse PruA (blue) with the corresponding region of *Mtb*-PruA (green). Residue numbers are those for *Mtb*-PruA. In the mouse P5CDH enzyme (and in *Tt*-P5CDH) glutamate binds in a pocket between the side chains of two Phe residues, Phe212 and Phe520. In contrast, in *Mtb*-PruA Phe498 (equivalent to Phe520) occupies the dimer interface as a result of a conformational change of residues 497–505. This leaves a much larger substrate-binding cavity in the *Mtb*-PruA active site, suggesting an attractive site for inhibitor design.

late stage of refinement, carried out further refinement with simulated annealing using *PHENIX* (Adams *et al.*, 2010) and rebuilt this part of the structure independently for each structure.

In addition to this specific difference in the GSA binding site, there is a difference in the structural response to NAD^+ binding. As noted above, the catalytic Cys327 must move aside to allow the correct binding of the nicotinamide ring. In *Tt*-P5CDH and *Mm*-P5CDH this simply involves a rotation about $C^{\alpha}-C^{\beta}$ and movement of S^{γ} by 3–4 Å. In *Mtb*-PruA, however, movement of Cys327 is associated with a substantial conformational change in the loop 326–329, altering the potential interactions that this loop can make in the active site.

3.9. Functional validation of the proline-utilization pathway

The substrate for *Mtb*-PruA, GSA (Fig. 1), is unstable and therefore has to be prepared *in situ* through the action of

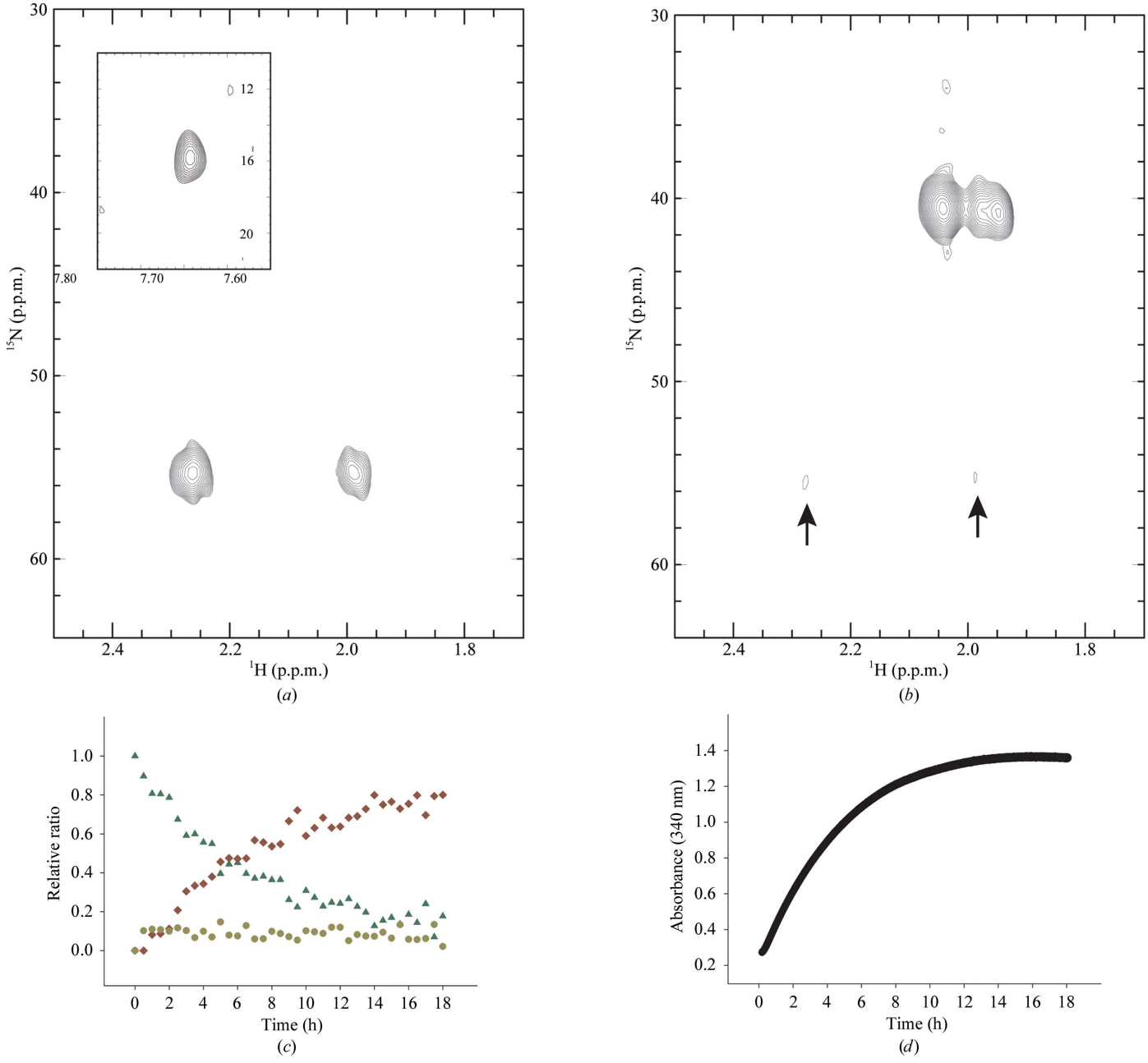


Figure 8
Functional characterization of the proline-utilization pathway in *Mtb*. Purified PruA and PruB enzymes were used to reconstitute the proline-utilization pathway and activity was then monitored by UV–visible and NMR spectroscopy. (*a*) The ^{15}N-proline signal from a two-dimensional ^{1}H–^{15}N NMR experiment at the start of the reaction. The inset indicates the signal detected from an intermediate of the reaction, believed to be P5C. (*b*) The ^{15}N-glutamate signal from the same experiment as (*a*) detected after 18 h. The arrows indicate residual signal from ^{15}N-proline at the end of the reaction. (*c*) The relative ratios of proline (red diamonds), glutamate (green triangles) and the intermediate (yellow circles) during the course of the reaction. (*d*) A duplicate reaction of (*c*) using UV–visible spectrophotometry, which monitors the production of glutamate from P5C by reduction of NAD^+ using PruA enzyme.

Mtb-PruB. In order to reconstitute the full proline-utilization pathway for *Mtb*, we therefore expressed and purified *Mtb*-PruB. Although the enzyme was soluble, it eluted in the void volume on size-exclusion chromatography, suggesting that the protein is in an aggregated form. It retained its FAD cofactor, however, as judged by its yellow colour and characteristic absorbance spectrum, implying that at least a fraction of the PruB molecules were folded. Extensive attempts to overcome this aggregation were unsuccessful, although both the solubility and purity were greatly increased by using BOG and glycerol at final concentrations of 1%(*w*/*v*) and 10%(*v*/*v*), respectively. Similar solubility issues were encountered in the recent functional study of *Mtb*-PruB (Serrano & Blanchard, 2013), where the authors used 0.1% Triton X-100 in the lysis buffer to obtain soluble protein.

Because of aggregation, a full kinetic analysis was not possible. Nevertheless, the overall conversion of proline to glutamate by PruB and PruA (Fig. 1) could be followed spectroscopically. UV–visible spectroscopy was used to find the optimum enzyme ratios and substrate and cofactor concentrations, monitoring the conversion of NAD^+ to NADH which accompanies the final, PruA-catalysed, step of the pathway. The optimal PruA:PruB ratio was determined to be 1:10 and this was then used in all functional experiments. With an initial concentration of 1 m*M* proline, the reaction showed an increase in the absorbance at 340 nm over the 18 h that the reaction was monitored (Fig. 8*d*). This unequivocally demonstrated that both enzymes are active and that PruB, although aggregated, produces enough P5C to provide the GSA substrate for conversion to glutamate by PruA.

The reaction was then further investigated by NMR spectroscopy. The α N atom of proline is contained within a pyrroline ring, making it an ideal situation to monitor proline consumption and glutamate production at the same time using ^{15}N NMR spectroscopy. The proline-utilization pathway was therefore reconstituted using the purified PruA and PruB enzymes as above, and its activity was investigated by two-dimensional ^{1}H–^{15}N NMR spectroscopy. The assay monitors ^{15}N-proline consumption and the subsequent production of ^{15}N-glutamate, with the signals from ^{15}N-labelled compounds being recorded every 30 min over 18 h. The intensities of the two singlet peaks corresponding to proline (2.2/54.7 and 1.9/54.9 p.p.m.; Fig. 8*a*) decreased throughout the reaction, concurrently with an accumulation of the glutamate signal (two singlets at 1.96/40.0 and 1.87/40.2 p.p.m.; Fig. 8*b*). The signals for both ^{15}N-labelled proline and glutamate assigned as H^{β} are the same as the signals from the respective controls (data not shown). An additional singlet appears at a relatively high proton chemical shift value (7.62/15.4 p.p.m.) with a significant signal-to-noise ratio (Fig. 8*a*, inset). We speculate that this singlet corresponds to the reaction intermediate P5C, which is produced from proline by PruB (Fig. 1). In accord with this conclusion, a signal at 7.78 p.p.m. has previously been assigned to H^{δ} in the P5C structure using ^{1}H-NMR (Farrant *et al.*, 2001). It is likely that the presence of the N=C$^{\delta}$ double bond in P5C prevents the exchange of H^{δ} with water molecules, allowing the H^{δ} signal to be detected. Whereas the intermediate level remains constant throughout the course of the reaction (Fig. 8*c*), almost all of the proline is converted to glutamate after 18 h, when only residual signal from proline is detectable (Figs. 8*b* and 8*c*).

3.10. Investigation of crystal disorder

A number of reports can be found in the literature describing crystal structures with similar arrangements of ordered and disordered layers, referred to as order–disorder and/or lattice-translocation defects (Wang, Kamtekar *et al.*, 2005; Wang, Rho *et al.*, 2005; Tsai *et al.*, 2009). We performed various analyses in order to find out the nature of the crystal disorder in the long-cell PruA crystals described here (Fig. 2). We considered the possibility of twinning complications, but twinning tests (Padilla & Yeates, 2003) provided by both *XDS* and *SCALA* showed that neither perfect nor partial merohedral twinning was present. We also noted that the diffraction patterns of the long-cell crystals showed an alternating pattern of reflections with strong and weak intensities along the *c* axis, suggesting the possibility of either a pseudo-translation in space group *P*622 or an additional screw axis. The data could be re-indexed in space group $P6_322$ by *POINTLESS*, but in this space group the structure could not be refined below R = 40%, whereas the final R factor in space group *P*622 is 25%. The native Patterson map for the long-cell apo PruA crystal had a peak at (u = 0.0145, v = 0.0076, w = 0.5) with a height corresponding to 16.8% of the origin-peak intensity. A similar peak, with height 15% of the origin peak, was found for the long-cell PruA–NAD^+ crystal data. No other significant non-origin peaks were found; the next highest was at 2.5% of the origin peak. This indicates that a pseudo-translation equivalent to half a unit cell exists along the *c* axis in both long-cell crystals and explains the pattern of strong and weak intensities. It also confirms *P*622 as the likely correct space group.

We next attempted to define the structure in the disordered layer by collecting MAD data from a long-cell crystal and looking for potential positions of Se atoms within the empty layer. Whereas *SHELXD* had identified the locations of all eight possible Se atoms of the single PruA molecule in the short-cell crystal form, it located a larger number of Se atoms (up to 30) in the long-cell crystal form, with occupancies varying from 1.0 to 0.17. The eight Se atoms with the highest occupancies (1.0 to 0.6) were in the ordered layer of the crystal and corresponded to those in the short-cell crystal. The remaining, low-occupancy, sites were in the disordered layer, providing evidence that there is protein present and that an anomalous signal from the disordered protein can be detected. Although these low-occupancy sites were barely above the noise level, eight of them were found to be related to the eight high-occupancy sites by a 180° rotation and a (0, 0, 1/2) translation, consistent with the pattern of alternating strong and weak intensities along the *c* axis. An electron-density map phased by all 30 sites showed that the selenium sites in the disordered layers were associated with 'donut-shaped' electron density for the Se atoms rather than conventional spherical shapes (Fig. 9). This 'donut-shaped' electron density for

the Se atoms is probably indicative of different orientations of the molecule within the disordered layer, with the electron density being an average of all orientations.

The electron-density map phased by the full number of Se atoms, including those found in the disordered layer, was then submitted to *MOLREP*, searching for two monomers in the asymmetric unit. Two molecules were built, one in the ordered layer and another in the disordered layer, related to each other by a 180° rotation. The electron density for the molecule in the disordered layer was weak and noisy, however, compared with that for the molecule in the ordered layer. *MOLREP* positioned the second molecule such that the Se atoms were in the middle of the 'donut-shaped' electron density in the disordered layer, and this molecule in the disordered layer could not be refined satisfactorily, although some helices appeared to be in the right positions. Comparison of the R and $R_{\rm free}$ values for the structures with one or two copies in the asymmetric unit showed little difference, suggesting that the structure in the missing layer contributes little to the overall diffraction.

The order–disorder phenomenon seen in our long-cell crystals appears to be different from previously described examples, in that it has not been possible to model the structure in the disordered layer, probably because the molecules have too many different orientations. Our results suggest that the PruA molecules adopt varying orientations within these layers, possibly through a relatively small-scale in-plane disorder. Other examples have been described with alternating ordered and disordered layers as in the present case (Wang, Kamtekar *et al.*, 2005; Wang, Rho *et al.*, 2005; Tsai *et al.*, 2009), but with a small number of orientations (three in the case of the carboxysome shell protein described by Tsai *et al.*, 2009). An alternative type of order–disorder phenomenon is that in which molecules are arranged with a defined difference in orientation in different unit cells, but randomly distributed through the crystal, not alternating (Pletnev *et al.*, 2009).

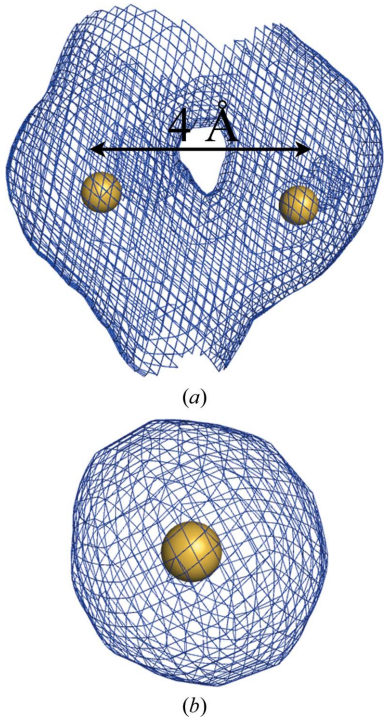


Figure 9
Electron density of the SeMet89 residue. (*a*) The 'donut-shaped' $2F_{\rm o} - F_{\rm c}$ electron density associated with SeMet89 in the empty layer; (*b*) indicates the electron density for the same residue in the ordered layer. Electron-density maps are contoured at 4.0 and 8.0σ for the disordered and ordered layer, respectively, without any refinement after phasing. The positions of Se atoms are shown by yellow spheres within the electron density.

It is not clear what the origins of the crystal disorder may be in the present case, although it may be significant that only in the ordered layer of the long-cell crystals was the catalytic Cys327 unmodified by β-ME and that attempts to reproduce crystals in the same condition without β-ME were not successful. Although we could not identify any link between crystal packing and the modification of Cys327 with β-ME, nor could we find any β-ME molecules at intermolecular sites, we speculate that β-ME is somehow implicated in the formation of the two crystal forms.

4. Conclusions

Proline metabolism has been shown to be an important player in the persistence of *Mtb* through detoxification of methylglyoxal (Berney *et al.*, 2012). In *Mtb*, proline utilization involves two monofunctional enzymes, PruB and PruA, which are part of an operon and are transcribed as a single mRNA (Berney *et al.*, 2012). Our structures of *Mtb*-PruA in its apo and NAD^+-bound states, although complicated by an order–disorder phenomenon in one crystal form, showed that the *Mtb* enzyme is very similar to other PruA enzymes, notably those from human and mouse, sharing most of the key structural elements involved in catalysis and binding. Interestingly, however, we find several specific conformational differences in the active site that were not anticipated from sequence comparisons, notably the displacement of a Phe residue that helps to define the binding site for GSA in the other P5CDH enzymes. These distinct differences, which appear to be specific to *Mtb*-PruA, could be exploited in drug design.

A feature of our study was the use of UV–visible and two-dimensional NMR spectroscopy to validate the predicted function of the entire two-enzyme pathway. By using NMR, we could directly monitor substrate consumption and product formation at the same time, without relying on indirect monitoring of the reactions. This is the first experimental characterization of the proline-utilization pathway in (micro)organisms containing monofunctional enzymes.

A remaining question is the possibility of molecular interaction between PruA and PruB, through which P5C may be passed from one enzyme to the other. Substrate channelling has been previously reported for the bifunctional PutA from *B. japonicum* (Srivastava *et al.*, 2010; Arentson *et al.*, 2012), in which the P5C intermediate is directly transferred from one domain to the other without being released into solution. We have been unable to find evidence for stable complex formation between PruA and PruB, however, despite co-expressing

these proteins under the control of single or separate promoters and fishing for complexes with one or other of the proteins affinity-tagged. Such an interaction seems possible, especially because the active site (Cys327 in particular) is solvent-accessible *via* a narrow channel that is potentially suitable for channelling substrate.

Atomic coordinates and structure-factor amplitudes have been deposited in the Protein Data Bank under accession codes 4ids, 4idm, 4ihi and 4jdc, corresponding to the native form (with gaps), the SeMet-substituted form, the NAD^+-bound form with the longer unit cell and the NAD^+-bound form with the shorter unit cell, respectively.

We thank Dr Michael Schmitz for help with NMR experiments and Drs David Goldstone and Christopher Squire for helpful discussions. TL is the recipient of a PhD scholarship from the Maurice Wilkins Centre for Molecular Biodiscovery. This research was funded by the Health Research Council of New Zealand. Access to the Australian Synchrotron was supported by the New Zealand Synchrotron Group Ltd.

References

Adams, E. & Frank, L. (1980). *Annu. Rev. Biochem.* **49**, 1005–1061.
Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Arentson, B. W., Sanyal, N. & Becker, D. F. (2012). *Front. Biosci.* **17**, 375–388.
Bashiri, G., Rehan, A. M., Greenwood, D. R., Dickson, J. M. & Baker, E. N. (2010). *PLoS One*, **5**, e15803.
Bashiri, G., Squire, C. J., Baker, E. N. & Moreland, N. J. (2007). *Protein Expr. Purif.* **54**, 38–44.
Berney, M. & Cook, G. M. (2010). *PLoS One*, **5**, e8614.
Berney, M., Weimar, M. R., Heikal, A. & Cook, G. M. (2012). *Mol. Microbiol.* **84**, 664–681.
Bottoms, C. A., Smith, P. E. & Tanner, J. J. (2002). *Protein Sci.* **11**, 2125–2137.
Bricogne, G. (1993). *Acta Cryst.* D**49**, 37–60.
Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). *J. Biomol. NMR*, **6**, 277–293.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Farrant, R. D., Walker, V., Mills, G. A., Mellor, J. M. & Langley, G. J. (2001). *J. Biol. Chem.* **276**, 15107–15116.
Gorrec, F. (2009). *J. Appl. Cryst.* **42**, 1035–1042.
Griffin, J. E., Gawronski, J. D., Dejesus, M. A., Ioerger, T. R., Akerley, B. J. & Sassetti, C. M. (2011). *PLoS Pathog.* **7**, e1002251.
Huang, T.-C., Huang, Y.-W., Hung, H.-J., Ho, C.-T. & Wu, M.-L. (2007). *J. Agr. Food Chem.* **55**, 5097–5102.
Huang, T.-C., Teng, C.-S., Chang, J.-L., Chuang, H.-S., Ho, C.-T. & Wu, M.-L. (2008). *J. Agr. Food Chem.* **56**, 7399–7404.
Inagaki, E., Ohshima, N., Takahashi, H., Kuroishi, C., Yokoyama, S. & Tahirov, T. H. (2006). *J. Mol. Biol.* **362**, 490–501.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kalapos, M. P. (1999). *Toxicol. Lett.* **110**, 145–175.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Lee, Y.-H., Nadaraia, S., Gu, D., Becker, D. F. & Tanner, J. J. (2003). *Nature Struct. Biol.* **10**, 109–114.
Luo, M., Arentson, B. W., Srivastava, D., Becker, D. F. & Tanner, J. J. (2012). *Biochemistry*, **51**, 10099–10108.
Moreland, N., Ashton, R., Baker, H. M., Ivanovic, I., Patterson, S., Arcus, V. L., Baker, E. N. & Lott, J. S. (2005). *Acta Cryst.* D**61**, 1378–1385.
Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). *Methods Enzymol.* **374**, 229–244.
Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.
Phang, J. M., Donald, S. P., Pandhare, J. & Liu, Y. (2008). *Amino Acids*, **35**, 681–690.
Pletnev, S., Morozova, K. S., Verkhusha, V. V. & Dauter, Z. (2009). *Acta Cryst.* D**65**, 906–912.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Serrano, H. & Blanchard, J. S. (2013). *Biochemistry*, **52**, 5009–5015.
Shah, N. S. *et al.* (2007). *Emerg. Infect. Dis.* **13**, 380–387.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Sophos, N. A. & Vasiliou, V. (2003). *Chem. Biol. Interact.* **143–144**, 5–22.
Srivastava, D., Schuermann, J. P., White, T. A., Krishnan, N., Sanyal, N., Hura, G. L., Tan, A., Henzl, M. T., Becker, D. F. & Tanner, J. J. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 2878–2883.
Srivastava, D., Singh, R. K., Moxley, M. A., Henzl, M. T., Becker, D. F. & Tanner, J. J. (2012). *J. Mol. Biol.* **420**, 176–189.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Tanner, J. J. (2008). *Amino Acids*, **35**, 719–730.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D67, 235–242.
Tsai, Y., Sawaya, M. R. & Yeates, T. O. (2009). *Acta Cryst.* D**65**, 980–988.
Tsuge, H., Kawakami, R., Sakuraba, H., Ago, H., Miyano, M., Aki, K., Katunuma, N. & Ohshima, T. (2005). *J. Biol. Chem.* **280**, 31045–31049.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. (2005). *Proteins*, **59**, 687–696.
Wang, F., Jain, P., Gulten, G., Liu, Z., Feng, Y., Ganesula, K., Motiwala, A. S., Ioerger, T. R., Alland, D., Vilchèze, C., Jacobs, W. R. Jr & Sacchettini, J. C. (2010). *Antimicrob. Agents Chemother.* **54**, 3776–3782.
Wang, J., Kamtekar, S., Berman, A. J. & Steitz, T. A. (2005). *Acta Cryst.* D**61**, 67–74.
Wang, J., Rho, S.-H., Park, H. H. & Eom, S. H. (2005). *Acta Cryst.* D**61**, 932–941.
Wayne, L. G. & Sohaskey, C. D. (2001). *Annu. Rev. Microbiol.* **55**, 139–163.
White, T. A., Krishnan, N., Becker, D. F. & Tanner, J. J. (2007). *J. Biol. Chem.* **282**, 14316–14327.
World Health Organization (2011). *Towards Universal Access to Diagnosis and Treatment of Multidrug-resistant and Extensively Drug-resistant Tuberculosis by 2015: WHO Progress Report 2011.* Geneva: World Health Organization. http://whqlibdoc.who.int/publications/2011/9789241501330_eng.pdf.
Zhang, Y. J., Ioerger, T. R., Huttenhower, C., Long, J. E., Sassetti, C. M., Sacchettini, J. C. & Rubin, E. J. (2012). *PLoS Pathog.* **8**, e1002946.