Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of pyrR (Rv1379) from *Mycobacterium tuberculosis*: a persistence gene and protein drug target

Katherine A. Kantardjieff,[a] Carolina Vasquez,[a] Peter Castro,[a] Nancy M. Warfel,[a] Beom-Seop Rho,[b] Timothy Lekin,[c] Chang-Yub Kim,[b] Brent W. Segelke,[c] Thomas C. Terwilliger[b] and Bernhard Rupp[c]*

[a]Department of Chemistry and Biochemistry and W. M. Keck Foundation Center for Molecular Structure, California State University Fullerton, Fullerton, CA 92834, USA, [b]Bioscience Division, MS M888, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, and [c]Biology and Biotechnology Research Program, L-448, University of California, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

Correspondence e-mail: br@llnl.gov

Received 27 September 2004
Accepted 21 December 2004

PDB Reference: pyrR, 1w30, r1w30sf.

The *Mycobacterium tuberculosis pyrR* gene (Rv1379) encodes a protein that regulates the expression of pyrimidine-nucleotide biosynthesis (*pyr*) genes in a UMP-dependent manner. Because pyrimidine biosynthesis is an essential step in the progression of TB, the gene product pyrR is an attractive antitubercular drug target. The 1.9 Å native structure of Mtb pyrR determined by the TB Structural Genomics Consortium facilities in trigonal space group $P3_121$ is reported, with unit-cell parameters $a = 66.64$, $c = 154.72$ Å at 120 K and two molecules in the asymmetric unit. The three-dimensional structure and residual uracil phosphoribosyltransferase activity point to a common PRTase ancestor for pyrR. However, while PRPP- and UMP-binding sites have been retained in Mtb pyrR, a distinct dimer interaction among subunits creates a deep positively charged cleft capable of binding *pyr* mRNA. *In silico* screening of pyrimidine-nucleoside analogs has revealed a number of potential lead compounds that, if bound to Mtb pyrR, could facilitate transcriptional attenuation, particularly cyclopentenyl nucleosides.

1. Introduction

The TB Structural Genomics Consortium (TBSGC) is one of the nine NIGMS-funded Protein Structure Initiative pilot projects and serves as a structural biology resource for the *Mycobacterium tuberculosis* (Mtb) research community (Terwilliger *et al.*, 2003). Consortium members can target proteins of interest and highly ranked targets are produced at consortium protein-production facilities. Proteins are then shipped to the crystallization facility at Lawrence Livermore National Laboratory (LLNL) for automated high-throughput crystallization (Rupp *et al.*, 2002) and data collection at the Advanced Light Source (ALS) in Berkeley (Snell *et al.*, 2004). Structures are determined by the facilities or the targeting consortium members. Coordinates are deposited and released in accordance with the NIH guidelines for Structural Genomics Pilot Projects.

TB is a re-emerging disease with an increasing prevalence of multi-drug-resistant strains (Ramaswamy & Musser, 1998) and a long-term goal of the TBSGC is to provide a foundation for structure-guided drug design. Protein targets of high priority are those which are essential or unique to the bacillus, particularly those which target the bacillus in the persistent stage (Stewart *et al.*, 2003). One such target is pyrR, a protein encoded by the *pyrR* gene (Rv1379), which regulates expression of *pyr* genes on the pyrimidine-nucleotide biosynthesis operon in many bacteria. All organisms require pyrimidine metabolism to supply building blocks for the synthesis of DNA, RNA and certain enzymes in central

metabolism. In all three domains of life, the *de novo* synthesis of pyrimidines is universally performed in a pathway of six enzymatic steps leading to the formation of UMP, which is then converted to UTP, CTP, dCTP and dTTP. In *M. tuberculosis*, five pyrimidine-biosynthesis genes (*pyrB*, *pyrC*, *carA*, *carB* and *pyrF*) and the regulatory protein pyrR are located on the *pyr* operon (Cole *et al.*, 1998). *carA* encodes the small subunit of carbamoyl phosphate synthase, which catalyzes the formation of carbamoyl phosphate required for both pyrimidine and arginine biosynthesis in *M. tuberculosis* (Cunin *et al.*, 1986; Martinussen *et al.*, 2001).

1.1. The *pyr* pathway as a target for antitubercular drugs

Significant differences exist in the details and regulatory mechanisms of the pyrimidine-biosynthesis pathway in bacteria and mammals (Mizrahi *et al.*, 2000), which can be exploited to develop specific antibiotics. In bacteria and archaea, pyrimidine-biosythesis reactions are performed by monofunctional enzymes, each arising from a separate gene. By contrast, in mammals pyrimidine biosynthesis involves substrate channeling and multifunctional enzymes, where the first three enzyme activities are linked on a single multifunctional cytosolic polypeptide chain coded for by a single gene and a single mRNA molecule. The fourth enzymatic step in mammals occurs on the inner membrane of mitochondria and the last two steps are catalyzed by a single polypeptide coded for by a single gene and a single mRNA molecule. Because multifunctional enzymes facilitate substrate channelling in the eukaryotic pyrimidine-biosynthesis pathway, substrates are not diluted and intermediates do not accumulate. In bacteria, however, accumulation of the pyrimidine-biosynthesis intermediate carbamoyl aspartate is toxic (Turnbough & Bochner, 1985). Upregulation of pyrR during hypoxic stress, characteristic of the environment in the granuloma harboring *M. tuberculosis* during persistence (Nyka, 1974; Cole *et al.*, 1998), has been reported. Suppression of the resulting transcriptional attenuation of the pyrimidine-biosynthesis pathway could thus target Mtb in the difficult to treat (Stewart *et al.*, 2003) persistent stage.

In contrast, in the early phase of Mtb infection specific supression of pyrimidine biosynthesis would be desirable. *De novo* pyrimidine biosynthesis is generally activated in proliferating cells in response to an increased demand for nucleotides needed for DNA synthesis (Sigoillot *et al.*, 2003). In both bacteria and mammals, flux through the pyrimidine-biosynthesis pathway is feedback-regulated at the first rate-limiting step (negatively by pyrimidine nucleotides and positively by PRPP), but in mammals this involves a complex network of both allosteric and phosphorylation control mechanisms (Sigoillot *et al.*, 2003). In bacteria, the second regulatory mechanism involving autogenous transcriptional termination and antitermination *via* pyrR is proposed to control expression of the operon (Ghim *et al.*, 1994; Turner *et al.*, 1994; Ghim *et al.*, 1999). Notwithstanding the challenges of modeling or predicting allosteric attenuation of transcription within the operon by virtual screening, the putative regulatory protein pyrR is an attractive target for antitubercular drug development, particularly as differences between mammalian and bacterial pathways are exploited. Indeed, selective antibacterials acting on enzymes in the pyrimidine pathway (Copeland *et al.*, 2000) have been discovered.

1.2. Mechanism of pyrR regulation

PyrR is is proposed to act by binding specific sequences on *pyr* mRNA to attenuate transcription within the operon. Because pyrR binding to mRNA is tighter in the presence of UMP or UTP (Tomchick *et al.*, 1998), elevated intracellular levels of these nucleotides act to attenuate transcription, thus reducing expression of downstream *pyr* genes. PyrR proteins are evolutionarily related to uracil phosphoribosyltransferases (UPRTases), as demonstrated by sequence and structural similarities, and several pyrR proteins have been shown to be bifunctional (Van de Casteele *et al.*, 1997; Turner *et al.*, 1998; Ghim *et al.*, 1999). *Bacillus subtilis* pyrR, in addition to attenuating transcription at three sites within the operon, has been shown to possess UPRTase activity (Turner *et al.*, 1998), although most of the UPRTase activity and thus uracil salvage in *B. subtilis* arises from a second UPRTase-encoding gene, *upp* (Martinussen *et al.*, 1995), which is also present in Mtb. Mtb pyrR reportedly has weak catalytic activity (Cole *et al.*, 1998; Fields & Switzer, 2004), while pyrR from *Lactococcus lactis* shows no catalytic activity (Martinussen *et al.*, 2001). Thus, catalytic activity does not appear to be an essential feature of pyrR primary regulatory function, but may be an evolutionary remnant.

The structures of four pyrR proteins (*B. subtilis*, *B. caldolyticus*, *M. tuberculosis*, *Thermus thermophilus*) and three UPRTases (*B. caldolyticus*, *Toxoplasma gondii*, *Thermotoga maritima*) have been determined (Schumacher *et al.*, 1998; Tomchick *et al.*, 1998; Kadziola *et al.*, 2002; Smith, 2004), some in complex with ligands. We present here a brief structural description of Mtb pyrR (PDB code 1w30). A comparison with the other pyrR and related UPRTase structures provides structural evidence for the absence of significant catalytic activity and suggests a molecular mechanism for binding *pyr* mRNA. Preliminary simulated docking studies indicate that certain classes of nucleoside analogs may be potential drug leads.

2. Experimental methods

2.1. Cloning and expression

A 0.6 kbp DNA fragment containing the *pyrR* gene (Rv1379; Swiss-Prot accession No. P71807) was amplified by PCR from *M. tuberculosis* H37Rv genomic DNA as the template using the following oligonucleotide primers: 5′-AGATATA**CATATG**GGTGCTGCGGGTGATGCCGC-3′ and 5′-AATTC**GGATCC**TCGCGAGATCACCACGCCG-TCA-3′. The bases in bold represent the *Nde*I and *Bam*HI sites, respectively. The amplified DNA fragment was digested with *Nde*I and *Bam*HI restriction enzymes and subcloned into the corresponding restriction sites in the pET28b vector which

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution bin (1.92–1.85 Å).

Data collection	
Space group	$P3_121$
Wavelength (Å)	1.000
Unit-cell parameters	
$a = b$ (Å)	66.64
c (Å)	154.72
Resolution (Å)	38.8–1.85
Unique reflections	34571 (3402)
Redundancy	4.4 (4.0)
Completeness	99.6 (99.8)
R_{sym}	0.071 (0.423)
$\langle I/\sigma(I)\rangle$	13.0 (2.4)
No. of molecules in AU	2
V_M (Matthews coefficient) (Å^3 Da^{-1})	2.3
Solvent content (%)	46.5
Refinement	
Free R value, random 5%	0.244 (0.290)
R value	0.206 (0.278)
R.m.s.d. bond length† (Å)	0.020
R.m.s.d. bond angle† (°)	1.798
Overall coordinate error‡ (Å)	0.160
RSCC (*Shake&wARP*)§	0.90
RSCC (*REFMAC*5)¶	0.96
Ramachandran appearance: residues in††	
Most favored region	275 (93.9%)
Additional allowed region	15 (5.1%)
Generously allowed region	3 (1.0%)
Disallowed region	0

† Deviations from restraint targets (Engh & Huber, 1991). ‡ Estimated standard uncertainty, Diffraction Precision Index (DPI) based on R_{free} (Cruickshank, 1999). § Real-space correlation coefficient, F_c map against averaged and weighted *Shake&wARP* map (Reddy *et al.*, 2003). ¶ Real-space correlation coefficient, F_o map against F_c map, as reported by *REFMAC*5 (Murshudov *et al.*, 1997). †† Regions as defined in *PROCHECK* (Laskowski *et al.*, 1993).

had been modified to provide a C-terminal six-His tag immediately following the *Bam*HI site. *Escherichia coli* BL21 (DE3) cells were transformed with the pyrR-modified pET28b/His vector and grown to exponential phase at 310 K in TB media containing 35 µg ml^{-1} kanamycin. Expression of *pyrR* was induced with 0.5 m*M* IPTG and cells were harvested and pelleted after growth for 21 h at 293 K and stored at 193 K.

2.2. PyrR purification

The cell pellet was resuspended in 10 ml buffer *A* (20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl) per gram of cells and sonicated at 283 K with 30 s pulses for 10 min. The cell suspension was centrifuged at 100 000*g* for 30 min and the supernatant was filtered through a 0.2 µm pore membrane and loaded onto a 5 ml Talon superflow affinity column equilibrated with buffer *A*. After washing with 50 ml buffer *A*, the His-tagged pyrR was eluted from the cobalt-affinity column using buffer *B* (20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl and 300 m*M* imidazole). The eluant was dialyzed against dialysis buffer (20 m*M* Tris pH 8.0, 100 m*M* NaCl, 10 m*M* β-mercaptoethanol) and purified by gel filtration on a Superdex-75 column (Amersham Pharmacia Biotech). The peak fractions (monitored at OD_{280}) were analyzed by SDS–PAGE and the pooled protein fractions were concentrated to 9.8 mg ml^{-1} using a Centriprep YM-3 (Millipore) and stored at 277 K. The protein concentration was measured by Bradford assay (Bradford, 1976) using standard bovine γ-globulin (Bio-Rad). The purity was determined to be >95% by SDS–PAGE and MALDI–TOF mass spectroscopy (Applied Biosystems). The protein was stored at 277 K and shipped to the TB consortium crystallization facility at LLNL (Rupp *et al.*, 2002).

2.3. Crystallization

Sitting drops (0.5 µl protein solution plus 0.5 µl well solution) were robotically set up in IntelliPlates (Rupp *et al.*, 2002) using the Crystool random-screening protocol (Segelke, 2001). Of 288 conditions tested, a crystallization cocktail containing 0.1 *M* imidazole–maleate buffer pH 7.5 with 26%(*w*/*v*) PEG MME 2K and 2.8% EDTA yielded diffraction-quality crystals (1.85 Å), which were flash-cooled in liquid nitrogen.

2.4. Data collection

A cryoprotected crystal on a Hampton loop pin was robotically mounted on ALS beamline 5.0.3 (Snell *et al.*, 2004). Data to 1.85 Å were collected using a wavelength of 1.000 Å, integrated using *HKL*2000 and scaled with *SCALEPACK* (Otwinowski & Minor, 1997) in trigonal space group $P3_121$ (No. 152), with unit-cell parameters $a = 66.64$, $c = 154.72$ Å at 120 K. Data-collection statistics are summarized in Table 1 and details are available from the PDB header (1w30). Successful molecular replacement established $P3_121$ as the correct selection from the enantiomorphic pair (No. 152 *versus* No. 154).

2.5. Structure determination

The structure was determined by molecular replacement using a homology model based on the structure of pyrR from *B. subtilis* as the probe (Tomchick *et al.*, 1998), which shares 49% sequence identity with Mtb pyrR (Fig. 1). *EPMR* (Kissinger *et al.*, 1999) was used in automated partial structure mode, searching for two molecules. The search for the first molecule in the asymmetric unit converged at a correlation coefficient (CC) of 0.31. Using partial structure factors for solution one, the search for the second molecule gave a CC of 0.50 (R = 0.49) after multi-segment rigid-body refinement of three segments in each of the two individual molecules.

2.6. Model building and refinement

To ensure effective phase-bias removal, the *Shake&wARP* procedure was used as implemented in the TB consortium map-improvement server (Reddy *et al.*, 2003). The model was iteratively built into the resulting sixfold-averaged $F_o \times \mathrm{FOM}/\varphi_{wt}$ maps using the program *Xfit* from the *XtalView* package (McRee, 1999). Several loop regions with significantly different conformations in the two molecules were excluded from the NCS restraints. After repeated cycles of water building and real-space refinement followed by restrained *REFMAC*5 maximum-likelihood refinement (Murshudov *et al.*, 1997), the final structure refined to R = 0.205 and R_{free} = 0.244 (5% of data omitted at random from test set).

Real-space correlation coefficient plots were calculated (Reddy *et al.*, 2003). Details of the refinement and data-collection statistics are tabulated in the header file of the deposited PDB entry and summarized in Table 1.

2.7. Quality assessment

PROCHECK (Laskowski *et al.*, 1993) and *WHAT_CHECK* (Hooft *et al.*, 1996) reports were created upon coordinate deposition and are available from the PDB. Ramachandran plot distribution, coordinate error, r.m.s.d. from target geometry values and real-space correlation are typical for a well refined 1.9 Å structure and are summarized in Table 1.

2.8. Simulated docking

To examine the capability of Mtb pyrR to bind substrates and act catalytically and to virtually screen for potential inhibitors of UMP-complex formation, flexible docking simulations were performed with *ICM-Pro* v.3.1.02 (Abagyan *et al.*, 1994; Schapira, Abagyan *et al.*, 2003; Schapira, Raaka *et al.*, 2003). To test the robustness of the docking procedure, crystal structures of UPRTase ligand complexes from *B. caldolyticus* (Kadziola *et al.*, 2002) and *T. gondii* (Schumacher *et al.*, 1998) were simulated first. Structural superposition between Mtb pyrR and ligand-bound structures and inferences from sequence alignment were used to initially define the receptor site. *ICM PocketFinder*, which detects cavities of sufficient size to bind 'drugable' molecules, was also used to detect and assess the character of the active site in Mtb pyrR. Next, limited libraries of pyrimidine-nucleoside analogs (Cysyk *et al.*, 1995; Tarantino *et al.*, 1999; Choi *et al.*, 2000; Chun *et al.*, 2000; Gumina *et al.*, 2001; Kumar *et al.*, 2001; Song *et al.*, 2001; Naimi *et al.*, 2003) were virtually screened against Mtb pyrR using

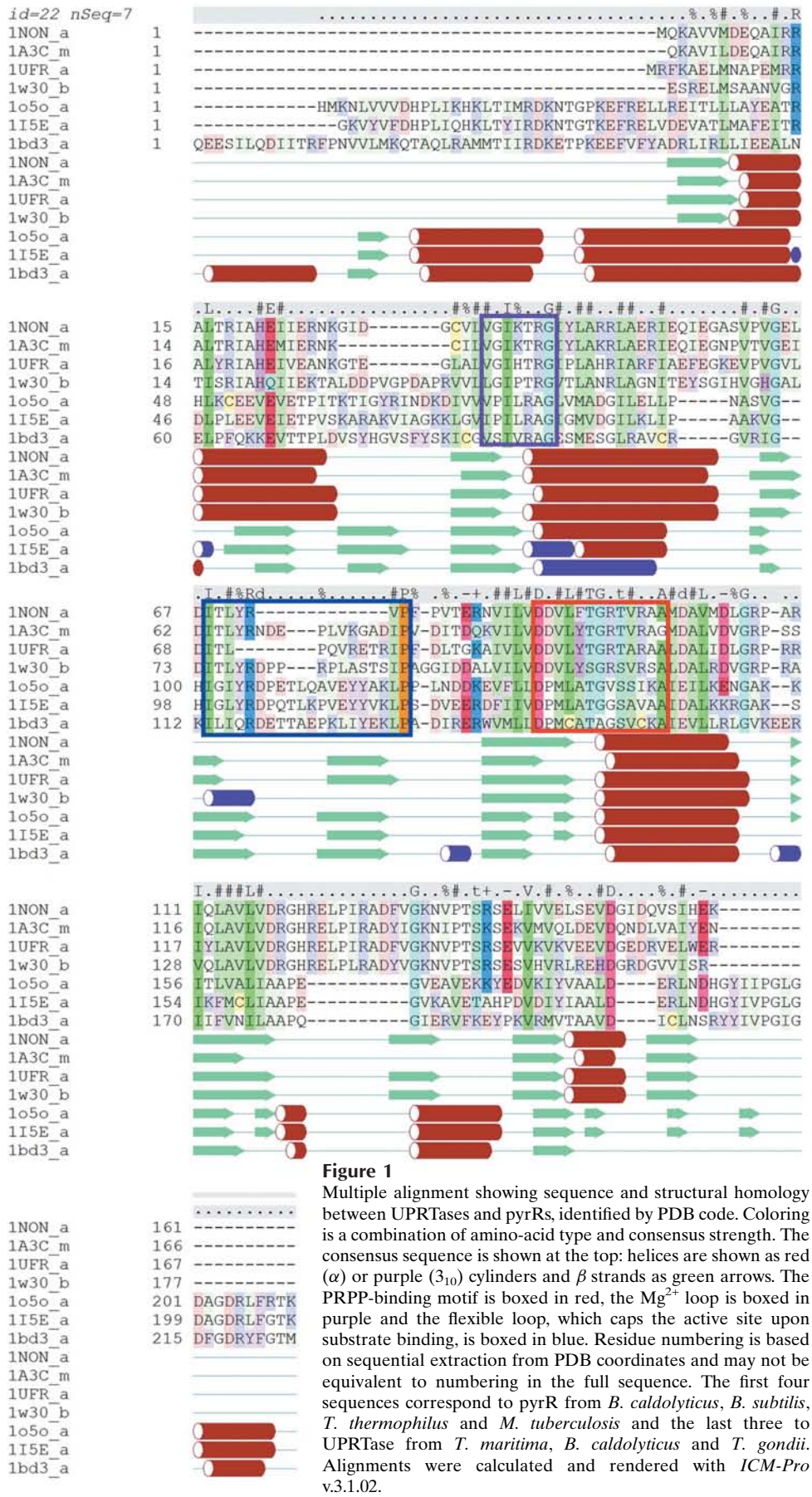

Figure 1
Multiple alignment showing sequence and structural homology between UPRTases and pyrRs, identified by PDB code. Coloring is a combination of amino-acid type and consensus strength. The consensus sequence is shown at the top: helices are shown as red (α) or purple (3_{10}) cylinders and β strands as green arrows. The PRPP-binding motif is boxed in red, the Mg^{2+} loop is boxed in purple and the flexible loop, which caps the active site upon substrate binding, is boxed in blue. Residue numbering is based on sequential extraction from PDB coordinates and may not be equivalent to numbering in the full sequence. The first four sequences correspond to pyrR from *B. caldolyticus*, *B. subtilis*, *T. thermophilus* and *M. tuberculosis* and the last three to UPRTase from *T. maritima*, *B. caldolyticus* and *T. gondii*. Alignments were calculated and rendered with *ICM-Pro* v.3.1.02.

ICM-Pro v.3.1.02 (Abagyan *et al.*, 1994; Schapira, Abagyan *et al.*, 2003; Schapira, Raaka *et al.*, 2003).

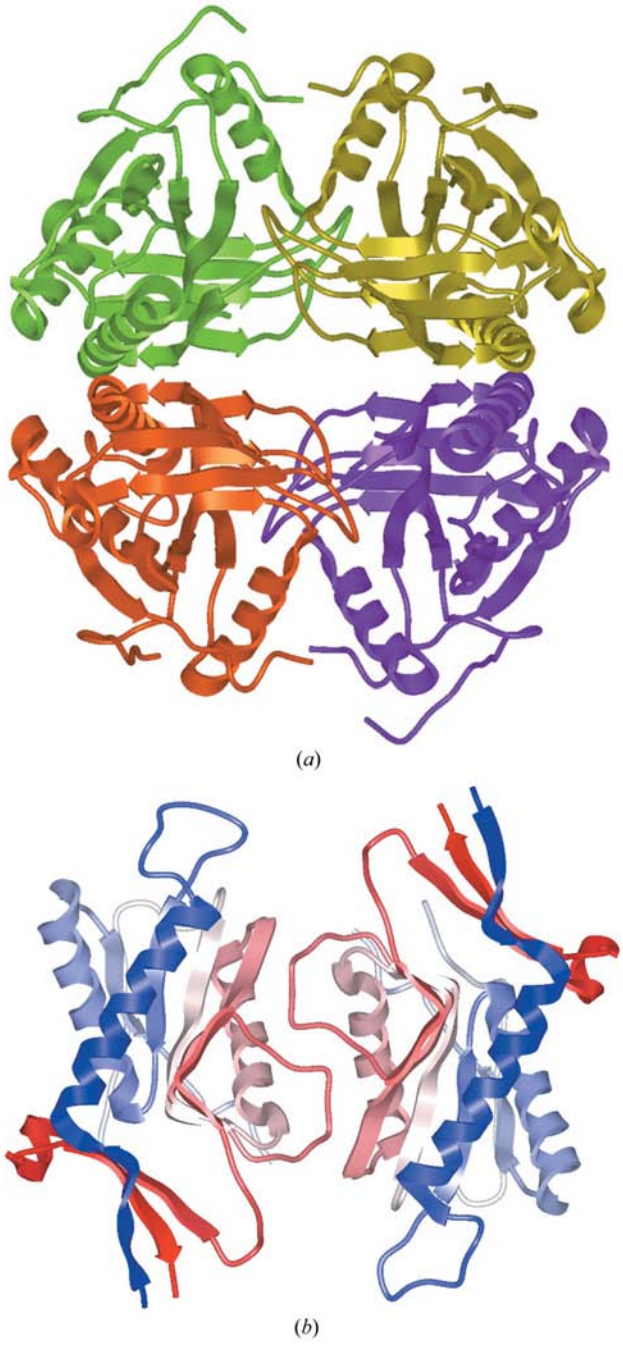

Figure 2
Mtb pyrR quaternary structure. (*a*) Mtb pyrR tetramer. The tetramer consists of a dimer of dimers. The homodimer is assembled from the green and yellow (or orange and purple) monomers of the figure. (*b*) Mtb homodimer, view rotated 90° into the page, with the orange/purple dimer in (*a*) removed from view. Each monomer consists of a 'core' parallel β-sheet domain and a secondary 'hood' domain. The core is reminiscent of the closely related UPRTases. Whereas the hood is predominantly α-helical in the UPRTases, the hood in pyrR terminates in a β-strand forming an antiparallel β-sheet with the C-terminus. PyrR subunits dimerize differently to the UPRTases. Coloring is blue to red from the N-terminus to the C-terminus. Rendered with *ICM-Pro* v.3.1.02.

3. Results

3.1. Structure summary

In the crystal form reported here, Mtb pyrR forms a tetramer consisting of a dimer of homodimers related by crystallographic symmetry (Fig. 2*a*). The molecules are arranged identically as seen in *B. caldolyticus* pyrR (PDB code 1non). Differences occur chiefly in loop conformations, notably the flexible loop (see §3.2), and in the conformation of β1 at the N-terminus. The Mtb pyrR homodimer (Fig. 2*b*) is identified by its similarity to physiologically relevant dimers identified in known pyrR structures (apo-structure PDB codes 1a3c, 1non and1ufr) and its greater buried surface area upon dimer formation of 920 Å^2. The alternative dimer (crystal packing) contacts bury 650 Å^2 of surface.

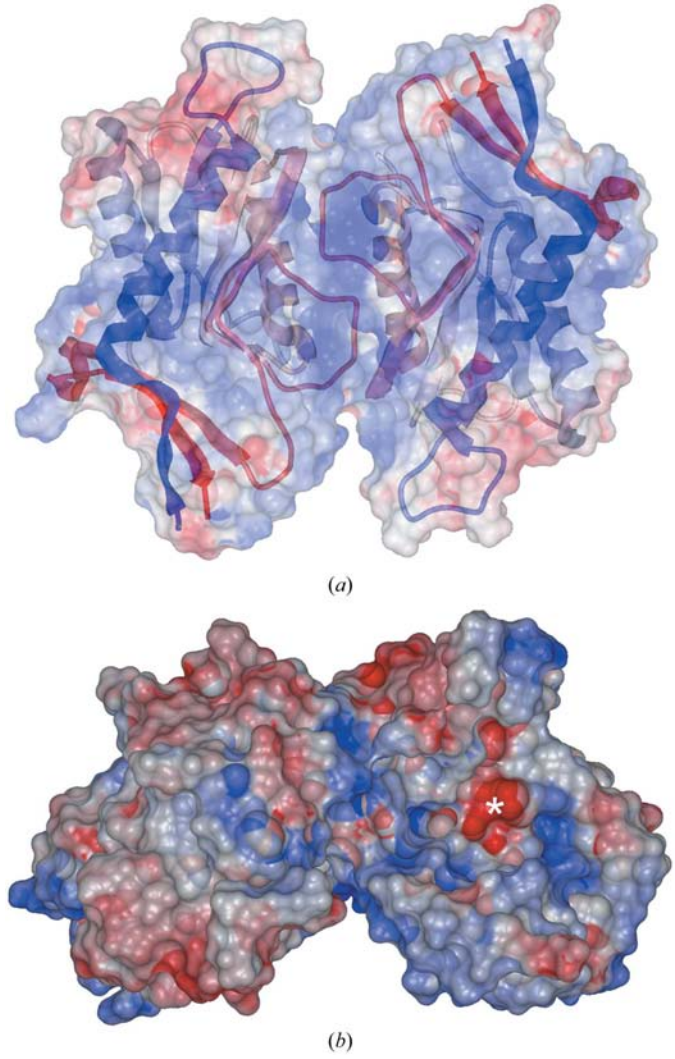

Figure 3
Electrostatic potential surfaces. (*a*) The electrostatic potential surface reveals that the Mtb pyrR homodimer possesses a deep positively charged cleft providing a molecular mechanism to bind *pyrR* mRNA, its primary function to regulate pyrimidine biosynthesis. (*b*) Orientation in (*a*) rotated 90° towards the viewer. The PRPP/UMP-binding region (*) is readily located by structural homology to the active site of UPRTases. Rendered with *ICM-Pro* v.3.1.02.

Mtb pyrR migrates by gel-filtration chromatography as though it has a molecular weight intermediate between dimeric and tetrameric forms. Given the relatively similar surface areas between dimer contacts and crystal contacts and the similar relatively small interface areas, anomalous migration behavior by gel filtration is not unexpected. The oligomerization state of pyrR is likely to be variable depending on concentration, pH and ionic strength (Tomchick *et al.*, 1998) and such anomalous migration behavior has been observed for other pyrRs (Turner *et al.*, 1998). While previous crystal structures have revealed dimeric, tetrameric and hexameric forms of pyrR, tetrameric and hexameric forms are believed to be non-physiological and pyrR is proposed to be a dimer at cellular concentrations (Tomchick *et al.*, 1998; Smith, 2004).

Each Mtb pyrR monomer consists of a 'core' parallel β-sheet domain and a secondary 'hood' domain. The core domain of Mtb pyrR is reminiscent of the closely related UPRTases (apo structure PDB codes 1bd3, 1i5e and 1o5o). Whereas the hood domain is predominantly α-helical in the UPRTases, the hood in Mtb pyrR, as seen in other pyrR structures (apo structure PDB codes 1w30, 1a3c, 1a4x, 1non and 1ufr), terminates in two β-strands which form an anti-parallel β-sheet with a third strand from the N-terminus. Although the recombinant Mtb pyrR protein possesses a C-terminal six-His tag, neither tag nor GS-linker residues are visible in the electron density.

The dimer interface between pyrR monomers in known structures is also different to that seen in the UPRTases. The Mtb pyrR electrostatic potential surface (Fig. 3*a*) reveals that the pyrR homodimer possesses a deep positively charged cleft on one face, which provides a molecular mechanism for binding *pyr* mRNA. Preliminary studies have established that Mtb pyrR binds the *M. tuberculosis* mRNA-binding loop in a UMP/UTP-dependent manner (Fields & Switzer, 2004). However, in the Mtb pyrR tetramer, as is seen in the tetrameric form of *B. caldolyticus* pyrR (PDB code 1non) and the hexameric form of *B. subtilis* pyrR (PDB code 1a4x), the basic surface faces an internal cavity which is not large enough to accommodate mRNA. For this reason, Mtb pyrR is likely to bind mRNA as a dimer.

The UPRTases have a conserved 13-residue fingerprint region, the PRPP-binding motif (Fig. 1), which is similar in MtB pyrRs and the PRPP-UMP binding pocket is easily located by structural homology to the active site of the UPRTases (Fig. 3*b*). While the solvent-accessible PRPP-binding regions in both monomers contain ligand electron density, this density is not interpretable beyond solvent molecules. The volume of the binding pocket in Mtb pyrR is 210 Å^3, significantly smaller than that found in the UPRTases (460–523 Å^3), but is comparable to other pyrRs (223–308 Å^3), and when clustered by binding pocket area and volume

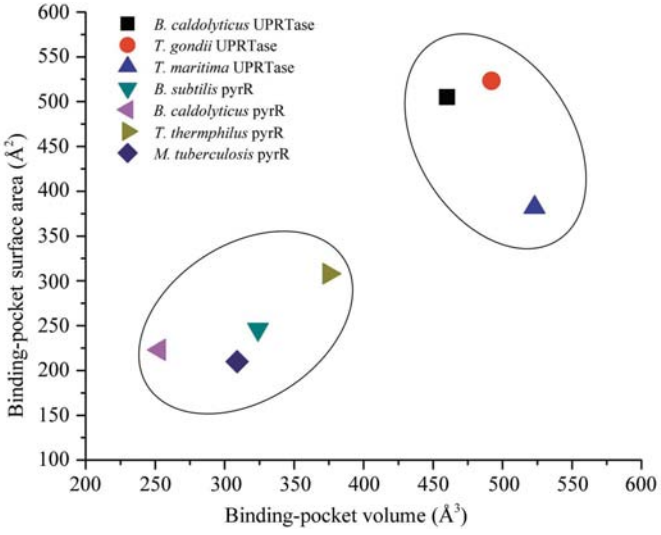

Figure 4
Comparision of binding-pocket characteristics in pyrRs and UPRTases. A plot of surface area *versus* volume of binding pockets, as calculated by *ICMPocketFinder* (Abagyan *et al.*, 1994; Schapira, Abagyan *et al.*, 2003; Schapira, Raaka *et al.*, 2003), shows that pyrRs and UPRTases fall into two distinct groups. PyrRs are located at the lower left of the plot and UPRTases at the upper right.

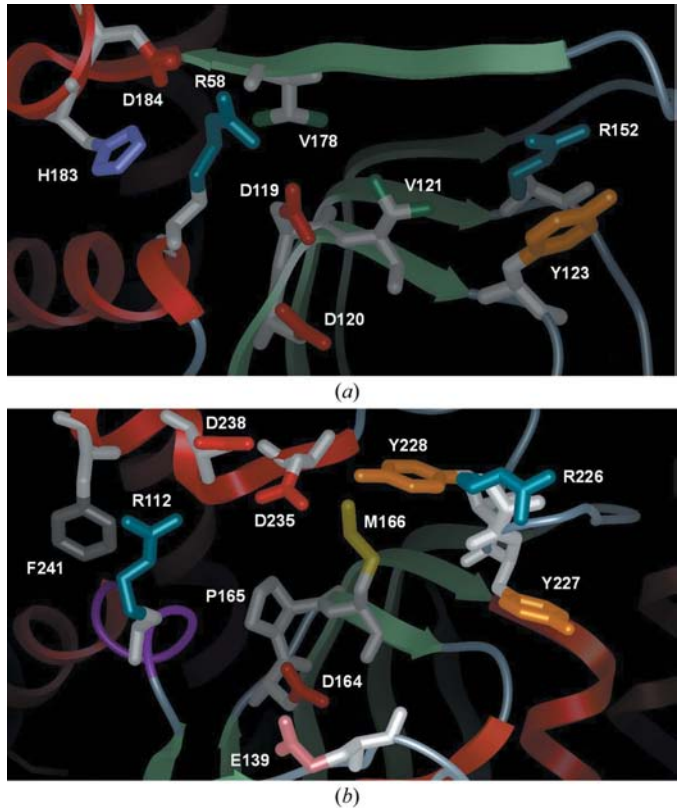

Figure 5
View into the binding pockets of Mtb pyrR and UPRTase from *T. gondii*. (*a*) The binding pocket of Mtb PyrR contains conserved residues in pyrRs and UPRTases, including a conserved aspartate (119) involved in binding the ribosyl moiety of PRPP and UMP. The C-terminal conserved aspartates (184, 187) proposed to stabilize the transition state (Kadziola *et al.*, 2002) are not structurally conserved and only Asp184 is visible in this view. The bulky hydrophobic residues in the hood that are characteristic of the uracil-binding region of the active site in proteins showing significant UPRTase activity are generally absent. (*b*) The binding pocket of *T. gondii* UPRTase contains hydrophobic residues in the hood domain, as well as aspartates that stabilize uracil binding and the transition state (Asp164, Asp235 and Asp238; Schumacher *et al.*, 1998). Residues are colored by type. Rendered with *ICM-Pro* v.3.1.02 (Abagyan *et al.*, 1994; Schapira, Abagyan *et al.*, 2003; Schapira, Raaka *et al.*, 2003).

UPRTases and pyrRs cluster into two distinct groups (Fig. 4). The binding pocket of Mtb PyrR contains many of the conserved residues in pyrRs and UPRTases, including a conserved aspartate (119) involved in binding the ribosyl moiety of PRPP and UMP. Although Mtb pyrR contains three sequentially conserved aspartate residues (119, 184 and 187), the latter two residues are not structurally conserved as they are located in the C-terminus of the hood (Fig. 1). These C-terminal aspartates are proposed to stabilize the transition state in UPRTases (Kadziola *et al.*, 2002). Mtb pyrR also lacks key bulky hydrophobic residues in the hood that are characteristic of the uracil-binding region of the active site in proteins showing significant UPRTase activity (Fig. 5).

3.2. Simulated docking

3.2.1. Catalytic mechanism. In the proposed UPRTase catalytic mechanism (Kadziola *et al.*, 2002), PRPP (5-phosphoribosyl-1-pyrophosphate) is bound to the enzyme, which requires Mg^{2+} for catalysis. Uracil enters the active site and is stabilized as the enol tautomer. Electron translocation takes place *via* an oxocarbonium intermediate, forming UMP. Sequential binding of PRPP and uracil to Mtb pyrR was simulated with *ICM-Pro* (Schapira, Abagyan *et al.*, 2003; Schapira, Raaka *et al.*, 2003) and high-scoring conformations were refined by explicit global optimization of surface side chains. This region of the Mtb pyrR sequence is identified in Fig. 1.

The predicted binding mode of UMP to Mtb pyrR (Fig. 6) is similar to that seen in the UPRTases (Kadziola *et al.*, 2002) and in a recently determined structure of *B. caldolyticus* pyrR (Smith, 2004), in which the phosphoribosyl moiety of UMP docks within the PRPP-binding motif, interacting with a conserved aspartate (119). PRPP is predicted to bind to Mtb pyrR within the PRPP-binding motif in a conformation analogous to that seen in the crystal structure of the UPRTase from *T. gondii* (Schumacher *et al.*, 1998). The predicted binding mode of uracil, as the enol tautomer, appears conducive to electron transfer, although the binding pocket of Mtb pyrR is shallow, leaving both substrates solvent-exposed (Fig. 7). As noted earlier, Mtb pyrR lacks key hydrophobic residues in the hood that are proposed to stabilize uracil binding for catalysis. Moreover, the sequentially conserved aspartate residues that are alleged to stabilize the transition state in UPRTases (Kadziola *et al.*, 2002) are not structurally conserved. These differences between Mtb pyrR and

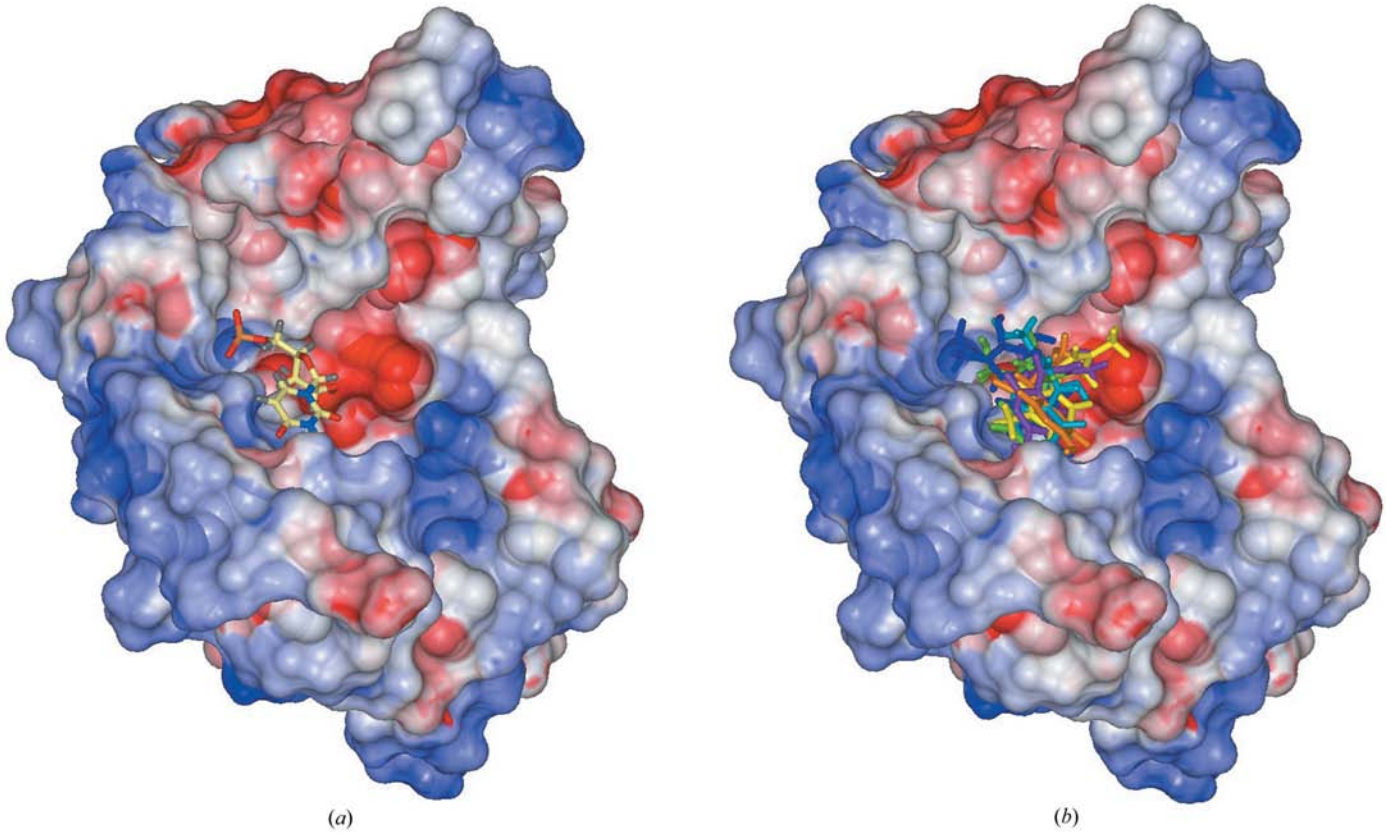

Figure 6
Virtual ligand screening against Mtb pyrR. (*a*) The predicted binding mode of UMP to Mtb pyrR is similar to that seen in the UPRTases. The binding pocket of Mtb PyrR contains many of the conserved residues in pyrRs and UPRTases, including a conserved aspartate (119) involved in binding the ribosyl moiety of PRPP and UMP. The orientation is the same as Fig. 2(*b*). (*b*) Limited libraries of pyrimidine-nucleoside analogs (Cysyk *et al.*, 1995; Tarantino *et al.*, 1999; Choi *et al.*, 2000; Chun *et al.*, 2000; Gumina *et al.*, 2001; Kumar *et al.*, 2001; Song *et al.*, 2001; Naimi *et al.*, 2003) screened against Mtb pyrR using *ICM-Pro* v.3.1.02 (Abagyan *et al.*, 1994; Schapira, Abagyan *et al.*, 2003; Schapira, Raaka *et al.*, 2003). Shown are a series of high scoring L- and D-cyclopentenyl nucleosides. Efforts are under way in our laboratory to determine crystal structures of Mtb pyrR in complex with several of these compounds, as well as to assay their *in vitro* and *in vivo* activity.

UPRTases may be generally extended to other pyrR structures in order to explain the differences in catalytic activity. An additional structural feature important to catalytic activity is a highly flexible loop that is not completely visible in any of the pyrR crystal structures (residues 91–98*A* and 91–95*B* in Mtb pyrR), which is believed to cap the active site during catalysis (Schumacher *et al.*, 1998). This loop appears to be structurally shorter in pyrRs than in UPRTases (Fig. 1). Homology modeling of the *L. lactis* pyrR (not shown), which has been reported to lack catalytic activity (Martinussen *et al.*, 2001), predicts a short flexible loop and a binding pocket with characteristics similar to the pyrRs noted earlier (Fig. 4).

An unresolved structural feature is the precise location of the catalytic magnesium ion in Mtb pyrR, which should be present for UPRTase activity. An approximate location of a magnesium ion-binding site in Mtb pyrR can be inferred by structural superposition with the UPRTase from *T. gondii* (Schumacher *et al.*, 1998), which reportedly contains a bound magnesium ion in one monomer. In *T. gondii* UPRTase, two sequentially conserved aspartates participate in Mg^{2+} binding (119 and 184 in Mtb pyrR). Only one of these residues is structurally conserved in pyrRs (119) and Arg58 in Mtb pyrR, which is highly conserved in pyrRs, occupies any possible metal-coordination site between aspartates. A catalytic magnesium ion would therefore have to bind differently to Mtb pyrR.

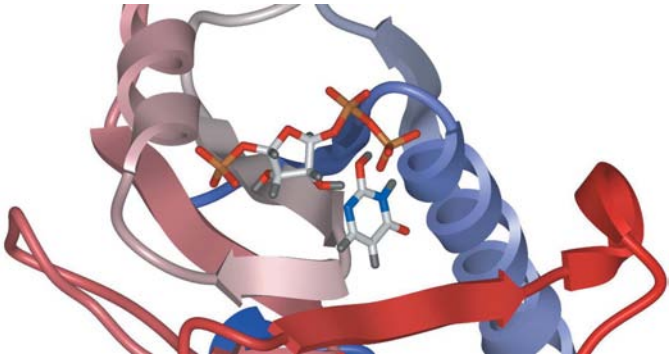

Figure 7
Simulated docking of PRPP and uracil to Mtb pyrR. The UPRTases have a conserved 13-residue fingerprint region, the PRPP-binding motif, which is similar in pyrRs and consists structurally of the β-strands and C-terminal helix, shown in pink (see also Fig. 1). Sequential binding of PRPP and uracil to Mtb pyrR was simulated with *ICM-Pro* v.3.1.02 and high-scoring conformations were refined by explicit global optimization of surface side chains (Abagyan *et al.*, 1994; Schapira, Abagyan *et al.*, 2003; Schapira, Raaka *et al.*, 2003). PRPP binds within the PRPP-binding motif in a conformation analogous to that seen in the crystal structure of the UPRTase from *T. gondii* (Schumacher *et al.*, 1998). The binding mode of uracil, as the enol tautomer, appears conducive to electron transfer. However, the binding pocket of Mtb pyrR is shallow, leaving both substrates solvent-exposed. Mtb pyrR lacks key hydrophobic residues in the C-terminal hood (red β-strand at bottom of figure; see also Fig. 5) that are proposed to stabilize uracil binding for catalysis. Moreover, sequentially conserved aspartate residues alleged to stabilize the transition state in UPRTases (Kadziola *et al.*, 2002; Asp184 and Asp187 located in the short helix and subsequent loop shown in red) are not structurally conserved. Coloring is blue to red from the N-terminus to C-terminus. Rendered with *ICM-Pro* v.3.1.02 (Abagyan *et al.*, 1994; Schapira, Abagyan *et al.*, 2003; Schapira, Raaka *et al.*, 2003).

Inspection of the PRPP/UMP-binding pockets of known pyrR structures suggests that residues implicated in Mg^{2+} binding could be located in proximity of the loop formed at the N-terminus of helix 2, with the sequence GI*X*T (residues 54–57 in Mtb pyrR; see Fig. 1). The residue at position *X*, which is usually Lys in pyrRs, would be capable of donating a backbone carbonyl O atom as a ligand to the metal. Additional putative ligands to the metal would be carboxyl O atoms from the conserved residues Asp119 and Asp120 at the N-terminus of the PRPP-binding motif (Fig. 1). Asp120, which is conserved in pyrR sequences, is Pro in UPRTases. This configuration of residues, where Lys, Asp and Asp ligate the metal, has been observed for known magnesium-binding sites in proteins (Feher *et al.*, 1998; Cates *et al.*, 2002; Mukhopadhyay *et al.*, 2004; Smith, 2004). In Mtb pyrR, residue *X* is Pro56, which creates a tighter turn at the helical cap and opens up the putative magnesium-binding site more to the solvent, possibly reducing affinity for a metal ion. Electron density in this region has been modeled as an ordered water molecule in both Mtb pyrR monomers. The density and bond distances to ligands Pro56, Asp119 and Asp120 are consistent with a water molecule and not a magnesium ion. Because the crystallization cocktail contained EDTA, which chelates magnesium, a crystal structure in the presence of magnesium and absence of EDTA remains to be determined to establish the precise location for magnesium binding to Mtb pyrR. Since the submission of this manuscript, a structure of *B. caldolyticus* pyrR has been determined that reveals magnesium bound in a manner similar to that we have proposed for Mtb pyrR (Smith, 2004).

3.2.2. Virtual screening. Virtual screening of limited libraries of pyrimidine-nucleoside analogs (Cysyk *et al.*, 1995; Tarantino *et al.*, 1999; Choi *et al.*, 2000; Chun *et al.*, 2000; Gumina *et al.*, 2001; Kumar *et al.*, 2001; Song *et al.*, 2001; Naimi *et al.*, 2003) against Mtb pyrR using *ICM-Pro* v.3.1.02 (Abagyan *et al.*, 1994; Schapira, Abagyan *et al.*, 2003; Schapira, Raaka *et al.*, 2003) has provided insights into structure–activity relationships that may be used to rationalize and prioritize compound library design (Burley, 2004). Analogs based on uracil, cytosine or thymine were predicted to bind to Mtb pyrR with affinity comparable to UMP. Cyclopentenyl nucleoside analogs scored particularly well, with L-nucleosides scoring slightly better than D-nucleosides (Fig. 6). Addition of functionalities larger than methyl to the 5-position of the ring generally decreased the binding affinity of the analog for the protein, as did halogenation (for example, 5-fluoro or 5-iodo analogs). Analogs of the ribosyl moiety possessing hydrogen-bond donors capable of interacting with the PRPP-binding motif had higher predicted affinity for Mtb pyrR. Based on the results of virtual screening, we speculate that nucleoside analogs with predicted higher binding affinity for Mtb pyrR would compete with UMP to attenuate transcription within the operon through an allosteric mechanism. However, because allostery is not easily modelled by virtual screening, it is still possible that the selected nucleotide analogs may prevent UMP binding without affecting mRNA affinity. In this case, pyrimidine biosynthesis would not be adequately down-

regulated during hypoxic stress, which could be equally deleterious to Mtb (Turnbough & Bochner, 1985). Crystal structures of pyrR in complex with selected analogs remain to be determined and competition mRNA-binding assays are anticipated.

4. Conclusions

The primary physiological function of Mtb pyrR is regulating the expression of the *pyr* operon. The three-dimensional structure and remaining UPRTase activity point to a common PRTase ancestor for pyrR. While PRPP- and UMP-binding sites have been retained in Mtb pyrR and other known pyrR structures, a distinct dimer interaction among subunits creates a deep positively charged cleft adapted to mRNA binding. The weak catalytic activity exhibited by Mtb pyrR can be explained by structural determinants in the active site. The binding pocket for substrates is more solvent-exposed in Mtb pyrR and lacks key hydrophobic residues that stabilize uracil binding for catalysis. Moreover, the conserved aspartate residues alleged to stabilize the transition state in UPRTases are not structurally conserved in Mtb pyrR. A more subtle structural difference is the location of a catalytic Mg^{2+}-binding region in Mtb pyrR, which may not be present, may not have high affinity for the ion or may not optimally position the ion for catalysis. Thus, although the formation of a pyrR–UMP complex is needed for mRNA binding, Mtb pyrR is not optimized to produce UMP catalytically and unlike the UPRTases it cannot be targeted to produce suicide nucleotides.

These differences in the size, shape and charge disposition of the binding pocket from the active site in UPRTases may however be exploited to design selective ligands of Mtb pyrR. Simulated docking of pyrimidine-nucleoside analogs reveals a number of potential drug leads that, if bound to Mtb pyrR, could act to attenuate transcription, particularly cyclopentenyl nucleosides. Interestingly, cyclopentenyl uracil has been shown to be an effective inhibitor or uridine salvage *in vivo* as a non-toxic inhibitor of uridine kinase in mouse (Cysyk *et al.*, 1995) and a number of cyclopentenyl nucleosides are in phase I/II clinical trials as potent antivirals (Gumina *et al.*, 2001). Pyrimidine-nucleoside analogs may also have activity against Gram-positive bacteria (Tarantino *et al.*, 1999). Moreover, L-nucleosides, which scored well in this study, in addition to having activity comparable to their D-counterparts, may have more favourable toxicological profiles and greater metabolic stability (Gumina *et al.*, 2001). Thus, an informatic structure-guided approach to elucidate structure–activity relationships has led to the discovery of potential new antimycobacterial therapeutics targeting Mtb pyrR, a protein implicated in disease persistence. The outcome of this project will be a set of critically evaluated pyrR structures and corresponding series of lead compounds, the latter which will be screened *in vitro* and *in vivo*.

We acknowledge Uhn Soo Cho, Susan Wachocki, Minyoung So and Min-Young Kim for technical assistance with cloning, protein expression and purification at the LANL TB consortium protein-production facility and Chris Fields at UIUC for helpful discussion. NMW completed a portion of this work in fulfilment of a required senior thesis project at Bonita High School in La Verne, CA, USA. KAK thanks the California State University Program for Education and Research in Biotechnology and the W. M. Keck Foundation for support of the Center for Molecular Structure. LLNL is operated by University of California for the US DOE under contract W-7405-ENG-48. This work was funded by NIH P50 GM62410 (TB Structural Genomics) grant.

References

Abagyan, R., Totrov, M. & Kuznetsov, D. (1994). *J. Comput. Chem.* **15**, 488–506.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–252.
Burley, S. K. (2004). *Mod. Drug Discov.* **7**, 53–56.
Cates, M. S., Teodoro, M. L. & Phillips, J. N. Jr (2002). *Biophys. J.* **82**, 1133–1146.
Choi, Y., Li, L., Grill, S., Gullen, E., Lee, C. S., Gumina, G., Tsujii, E., Cheng, Y.-C. & Chu, C. K. (2000). *J. Med. Chem.* **43**, 2538–2546.
Chun, B. K., Olgen, S., Hong, J. H., Newton, G. & Chu, C. K. (2000). *J. Org. Chem.* **65**, 685–693.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
Copeland, R. A., Marcinkeviciene, J., Haquet, T. S., Kopcho, L. M., Jiang, W., Wang, K., Ecret, L. D., Sizemore, C., Amsler, K. A., Foster, L., Tadesse, S., Combs, A. P., Stern, A. M., Trainor, G. L., Slee, A., Rogers, M. J. & Hobbs, F. (2000). *J. Biol. Chem.* **275**, 33373–33378.
Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.
Cunin, R., Glansdorff, N., Pierard, A. & Stalon, V. (1986). *Microbiol. Rev.* **50**, 314–352.
Cysyk, R. L., Malinowski, N., Marquez, V. M., Zaharevitz, D., August, E. M. & Moyer, J. D. (1995). *Biochem. Pharmacol.* **49**, 203–207.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Feher, V. A., Tzeng, Y.-L., Hoch, J. A. & Cavanagh, J. (1998). *FEBS Lett.* **425**, 1–6.
Fields, C. J. & Switzer, R. L. (2004). Personal communication.
Ghim, S.-Y., Kim, C. C., Bonner, E. R., D'Elia, J. N., Grabner, G. K. & Switzer, R. L. (1999). *J. Bacteriol.* **181**, 1324–1329.
Ghim, S.-Y., Nielsen, P. & Neuhard, J. (1994). *Microbiology*, **140**, 479–491.
Gumina, G., Song, G.-Y. & Chu, C. K. (2001). *FEMS Microbiol. Lett.* **202**, 9–15.
Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.
Kadziola, A., Neuhard, J. & Larsen, S. (2002). *Acta Cryst.* D**58**, 936–945.
Kissinger, C. R., Gelhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Kumar, R., Rai, D., Sharma, S. K., Saffran, H. A., Blush, R. & Tyrrell, D. L. (2001). *J. Med. Chem.* **44**, 3531–3538.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Martinussen, J., Glaser, P., Anderson, P. S. & Saxild, H. H. (1995). *J. Bacteriol.* **177**, 271–274.
Martinussen, J., Schallert, J., Andersen, B. & Hammer, K. (2001). *J. Bacteriol.* **183**, 2785–2794.
Mizrahi, V., Dawes, S. S. & Rubin, H. (2000). *Molecular Genetics of Mycobacteria*, edited by G. F. Hatfull & W. R. Jacobs Jr, pp. 159–172. Washington DC, USA: ASM Press.
Mukhopadhyay, D., Sen, U., Zapf, J. & Varughese, K. I. (2004). *Acta Cryst.* D**60**, 638–645.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Naimi, E., Zhou, A., Khalili, P., Wiebe, L. I., Balzarini, J., de Clercq, E. & Knaus, E. E. (2003). *J. Med. Chem.* **46**, 995–1004.
Nyka, W. (1974). *Infect. Immun.* **9**, 843–850.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **267**, 307–326.
Ramaswamy, S. & Musser, J. M. (1998). *Tubercle Lung Dis.* **79**, 3–29.
Reddy, V., Swanson, S., Sacchettini, J. C., Kantardjieff, K. A., Segelke, B. & Rupp, B. (2003). *Acta Cryst.* D**59**, 2200–2210.
Rupp, B., Segelke, B. W., Krupka, H. I., Lekin, T. P., Schafer, J., Zemla, A., Toppani, D., Snell, G. & Earnest, T. E. (2002). *Acta Cryst.* D**58**, 1514–1518.
Schapira, M., Abagyan, R. & Totrov, M. (2003). *J. Med. Chem.* **46**, 3045–3059.
Schapira, M., Raaka, B. M., Das, S., Fan, L., Totrov, M., Zhou, Z., Wilson, S. R., Abagyan, R. & Samuels, H. H. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 7354–7359.
Schumacher, M. A., Carter, D., Scott, D. M., Roos, D. S., Ullman, B. & Brennan, R. G. (1998). *EMBO J.* **17**, 3219–3232.
Segelke, B. W. (2001). *J. Cryst. Growth*, **232**, 553–562.
Sigoillot, F. D., Berkowski, J. A., Sigoillot, S. M., Kotsis, D. H. & Guy, H. I. (2003). *J. Biol. Chem.* **278**, 3403–3409.
Smith, J. L. (2004). Personal communication.
Snell, G., Nordmeyer, R., Cornell, E., Meigs, G., Cork, C., Yegian, D., Jaklevic, J., Jin, J. & Earnest, T. E. (2004). *Structure*, **12**, 537–545.
Song, G.-Y., Paul, V., Choo, H., Morrey, J., Sidwell, R. W., Schinazi, R. F. & Chu, C. K. (2001). *J. Med. Chem.* **44**, 3985–3993.
Stewart, G. R., Robertson, B. D. & Young, D. B. (2003). *Nature Rev. Microbiol.* **1**, 97–105.
Tarantino, P. M., Zhi, C., Gambino, J. J., Wright, G. E. & Brown, N. C. (1999). *J. Med. Chem.* **42**, 2035–2040.
Terwilliger, T. C. *et al.* (2003). *Tuberculosis*, **83**, 223–249.
Tomchick, D. R., Turner, R. J., Switzer, R. L. & Smith, J. L. (1998). *Structure*, **6**, 337–350.
Turnbough, C. L. Jr & Bochner, B. R. (1985). *J. Bacteriol.* **163**, 500–505.
Turner, R. J., Bonner, E. R., Grabner, G. K. & Switzer, R. L. (1998). *J. Biol. Chem.* **273**, 5932–5938.
Turner, R. J., Lu, Y. & Switzer, R. L. (1994). *J. Bacteriol.* **176**, 3708–3722.
Van de Casteele, M., Chen, P., Roovers, M., Legrain, C. & Glansdorff, N. (1997). *J. Bacteriol.* **179**, 3470–3481.