LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-207072

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

K. A. Kantardjieff, C. Vasquez, P. Castro, N. M. Warfel, B.-S. Rho, T. Lekin, C.-Y. Kim, B. W. Segelke, T. C. Terwilliger, B. Rupp

October 6, 2004

Acta Crystallographica D

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. E-mail: br@llnl.gov

Synopsis The 1.9 Å native structure of pyrimidine biosynthesis regulatory protein encoded by the *Mycobacterium tuberculosis pyrR* gene (Rv1379) is reported. Because pyrmidine biosynthesis is an essential step in the progression of TB, pyrR is an attractive antitubercular drug target.

Abstract The *Mycobacterium tuberculosis pyrR* gene (Rv1379) encodes a protein that regulates 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. We report the 1.9 Å native structure of Mtb pyrR determined by the TB Structural Genomics Consortium facilities (PDB entry 1W30) in trigonal space group $P3_121$, with cell dimensions at 120K of a = 66.64 Å, c = 154.72 Å, and two molecules in the asymmetric unit. The 3D 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 novel 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 leads compounds that, if bound to Mtb pyrR, could facilitate transcriptional attenuation, particularly cyclopentenyl nucleosides.

Keywords: Mycobacterium tuberculosis; pyr, Rv1379; pyrimidine biosynthesis; RNA-binding; drug target structure

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 immediately in accordance with NIH guidelines for Structural Genomics Pilot Projects..

TB is a re-emerging disease with 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. One such target is pyrR, a protein that regulates expression of genes and operons of pyrimidine nucleotide biosynthesis (*pyr* genes) in many bacteria. The *de novo* synthesis of pyrimidines is universal, 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*, six pyrimidine biosynthesis genes (*pyrR*, *pyrB*, *pyrC*, *carA*, *carB* and *pyrF*) 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 (Cunin et al. 1986; Martinussen et al. 2001). PyrR, encoded by the *pyrR* gene (Rv1379) is a putative regulatory protein, believed to act by binding specific sequences on *pyr* mRNA. 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. It is assumed that the pyrimidine biosynthesis pathway is similar to the well-established pathway in *E. coli* and mammalian cells, although unique regulatory properties are expected (Mizrahi et al. 2000).

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). *B. 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*, is due to a second UPRTase-encoding gene, *upp* (Martinussen et al. 1995), which is also present in Mtb (Cole et al. 1998). Mtb pyrR

reportedly has weak catalytic activity (Cole et al. 1998), while pyrR from *L. 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.

Mtb pyrR has been shown to be upregulated during hypoxic stress, characteristic of the environment found in the granuloma harboring *M. tuberculosis* (Nyka 1974; Cole et al. 1998). Moreover, deletion of the *pyrR* allele in *B. subtilis* produces resistance to the toxic pyrimidine analog 5-fluorouracil (Martinussen et al. 2001). Thus, our objectives in this work are to understand the structural relationships among pyrRs and the related UPRTases, and how pyrR recognizes specific *pyr* mRNA sequences in a UMP-, UTP- dependent manner. Moreover, because pyrimidine biosynthesis is an essential step in the progression of tuberculosis (Chan et al. 2002), Mtb *pyrR* is an attractive antitubercular drug target.

The structures of four pyrR proteins (*B. subtilis, B. caldolyticus, M. tuberculosis, T. thermophilus*) and three UPRTases (*B. caldolyticus, T. gondii, T. 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 structure description of Mtb pyrR (PDB entry 1W30) and a comparison to the other pyrR and related UPRTase structures, providing structural evidence for the absence of significant catalytic activity. Results from preliminary simulated docking studies suggest certain classes of nucleoside analogs may be potential drug leads.

2. Experimental Methods

2.1. Cloning and Expression

A 0.6 kb DNA fragment containing the *pyrR* gene (Rv1379, Swiss Prot accession number P71807), was amplified by PCR from *M. tuberculosis* H37Rv genomic DNA as the template, using the following oligonucleotide primers: 5'- AGATATA CATATGGGTGCTGCGGGTGA TGCCGC-3' and 5'-AATTCGGATCC TCGCGAGATCACCACGCCGTCA-3'. The underlined bases represent the NdeI and BamHI sites, respectively. The amplified DNA fragment was digested with NdeI and BamHI restriction enzymes, and subcloned into the corresponding restriction sites in the pET28b vector which had been modified to provide a C-terminal 6-His tag immediately following the BamHI site. *E. coli* BL21 (DE3) cells were transformed with the pyrR-modified pET28b/His vector and grown to exponential phase at 37ºC in TB media containing kanamycin. Expression of *pyrR* was induced with 0.5mM IPTG, and cells were harvested and pelleted after growth for 21 hours at 20ºC and stored at - 80ºC.

2.2. PyrR Purification

The cell pellet was resuspended in 10ml buffer A (20mM Tris-HCl, pH 8.0, 100mM NaCl) per gram of cells and sonicated at 10ºC with 30 second pulses for 10 min. The cell suspension was centrifuged at 100,000g for 30 min, and supernatant was filtered through a 0.2 µm pore membrane and loaded on 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 mM Tris-HCl, pH8.0, 500 mM NaCl and 300 mM imidazole). Eluant was dialyzed against dialysis buffer (20 mM Tris pH 8.0, 100 mM NaCl, 10 mM β–mercaptoethanol) and purified by gel filtration on a Superdex-75 column (Amersham Pharmacia Biotech). Peak fractions (monitored at OD280) were analyzed by SDS-PAGE and the pooled protein fractions were concentrated to 9.8 mg/ml using a Centriprep YM-3 (Millipore) and stored at 4ºC. Purity was determined to be >95% by SDS PAGE and MALDI-TOF mass spectroscopy (Applied Biosystems), was stored at 4ºC and shipped to the TB consortium crystallization facility at LLNL (Rupp et al. 2002).

2.3. Crystallization

Sitting drops (0.5uL protein solution + 0.5uL 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 (1.85Å) which were flash cooled in liquid nitrogen.

2.4. Data Collection

A cryo-protected crystal on a Hampton loop pin was robotically mounted on ALS beam line 5.0.3 (Snell et al. 2004). Data to 1.85 Å were collected at 1.000Å, integrated using HKL2000, and scaled with ScalePack (Otwinowski & Minor 1997) in trigonal space group $P3_121$ (no.152), with cell dimensions at 120K of a = 66.74 Å, c = 154.72 Å. Data collection statistics are summarized in Table 1, and details are available from the PDB header (1W30). Successful molecular replacement established $P\ 3_121$ as the correct selection from the enantiomorphic pair (no. 152 vs. no. 154).

2.5. Structure Determination

The structure was determined by molecular replacement using a homology model based on the structure of pyrR from *Bacillus subtilis* as the probe (Tomchick et al. 1998), which shares 49% sequence identity with Mtb pyrR (Figure 1). *Epmr* (Kissinger et al. 1999) was used in

automated partial structure mode searching for two molecules. Search for the first molecule in the asymmetric unit converged at 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, as implemented in the TB consortium map improvement server (Reddy et al. 2003), was used. The model was iteratively built into the resulting six-fold averaged Fo*f.o.m/PHIwt maps using the program *Xfit* in 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 *Refmac5* maximum likelihood refinement (Murshudov et al. 1997) the final structure (1W30) refined to R = 0.205 and freeR = 0.244. 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* (Hoft et al. 1996) reports were created upon coordinate deposition and are available from the PDB for entry 1W30. 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 3.1.02 (Abagyan et al. 1994; Schapira et al. 2003a; Schapira et al. 2003b). 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 ICM-Pro 3.1.02 (Abagyan et al. 1994; Schapira et al. 2003a; Schapira et al. 2003b).

3. Results

3.1. Structure Summary

Mtb pyrR is a homodimer (Figure 2), with a buried surface area upon dimer formation of 750Å^2 per monomer. Each monomer consists of a "core" parallel beta-sheet domain and a secondary "hood" domain. The core domain of Mtb pyrR is reminiscent of the closely related UPRTases (apo- structure PDB IDs: 1BD3, 1I5E, 1O5O). Whereas the hood domain is predominantly alpha helical in the UPRTases, the hood in Mtb pyrR, as seen in other pyrR structures (apo-structure PDB IDs: 1W30, 1A3C, 1NON, 1UFR) terminates in two beta strands, which form an antiparallel beta sheet with a third strand from the N-terminus. Although the recombinant Mtb pyrR protein possesses a C-terminal 6-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 than seen in the UPRTases. The Mtb pyrR electrostatic potential surface (Figure 3) reveals that the pyrR homodimer possesses a deep positively charged cleft adapted to mRNA binding.

The UPRTases have a conserved 13-residue fingerprint region, the PRPP binding motif (Figure 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. 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 210Å^3, significantly smaller than that found in the UPRTases (460 – 523Å^3), but comparable to other pyrRs (223 – 308Å^3), and when clustered by binding pocket area and volume, UPRTases and pyrRs cluster into two distinct groups (Figure 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, 187), the latter two residues are not structurally conserved, as they are located in the C-terminus of the hood (Figure 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, characteristic of the uracil binding region of the active site in proteins showing significant UPRTase activity (Figure 5).

3.2. Simulated Docking

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 et al. 2003a; Schapira et al. 2003b) and high-scoring conformations were refined by explicit global optimization of surface side-chains. This region of the Mtb pyrR sequence is identified in Figures 1.

The predicted binding mode of UMP to Mtb pyrR (Figure 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), where 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 (Figure 7). As noted earlier, Mtb pyrR lacks key hydrophobic residues in the hood which 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), are not structurally conserved. These differences between Mtb pyrR and UPRTases may be generally extended to other the pyrR structures to explain differences in catalytic activity. An additional structural feature important to catalytic activity is a highly flexible loop 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 (Figure 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 (Figure 4).

An unresolved structural feature is the location of the catalytic magnesium ion in Mtb pyrR. 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 R58 in Mtb pyrR,

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.

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 GIXT (residues 54-57 in Mtb pyrR; see Figure 1). The residue at position X, usually K in pyrRs, would be capable of donating a backbone carbonyl oxygen as a ligand to the metal. Additional putative ligands to the metal would be carboxyl oxygens from conserved residues D119 and D120 at the N-terminus of the PRPP-binding motif (Figure 1). D120, which is conserved among pyrR sequences, is P in UPRTases. This configuration of residues, where K, D and D ligate the metal, has been observed for known magnesium binding sites in proteins (Feher et al. 1998; Cates et al. 2002; Mukhapadhyay et al. 2004; Smith 2004). In Mtb pyrR, residue X is P56, 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 P56, D119 and D120 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 magnesium binding to Mtb pyrR.

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 3.1.02 (Abagyan et al. 1994; Schapira et al. 2003a; Schapira et al. 2003b) has provided insights into structure-activity relationships, which 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 (Figure 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.

4. Conclusions

The primary physiological function of Mtb pyrR is regulating expression of the *pyr* operon. The 3D 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 novel 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 stablize uracil binding for catalysis. Moreover, 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, 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 attentuate 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 phase1/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*.

Acknowledgements 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. 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) center grant.

References

Abagyan, R., Totrov, M. & Kuznetsov, D. (1994). *J Comp Chem* **15**, 488-506.

Burley, S.K. (2004). *Mod Drug Disc* **7**, 53-56.

Cates, M.S., Teodoro, M.L. & Phillips, J.N.J. (2002). *Biophys J* **82**, 1133-1146.

Chan, K., Knaak, T., Satkamp, L., Humbert, O., Falkow, S. & Ramakrishnan, L. (2002). *Proc Natl Acad Sci USA* **99**, 3920-3925.

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., Brosch, R., Parkill, J. & authors, o. (1998). *Nature* **393**, 537-544.

Cruickshank, D.W.J. (1999). *Acta Cryst.* **D55**, 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.* **A47**, 392-400.

Feher, V.A., Tzeng, Y.-L., Hoch, J.A. & Cavanagh, J. (1998). *FEBS Lett* **425**, 1-6.

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.

Gumina, G., Song, G.-Y. & Chu, C.K. (2001). *FEMS Microbiol Lett* **202**, 9-15.

Hoft, R.R.W., Vriend, G., Sander, C. & Albola, E.E. (1996). *Nature* **381**, 272-272.

Kadziola, A., Neuhard, J. & Larsen, S. (2002). *Acta Crystallog* **D**, 936-945.

Kissinger, C.R., Gelhaar, D.K. & Fogel, D.B. (1999). *Acta Crystallogr* **D55**, 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.

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.

McRee, D.E. (1999). *J Struct Biol* **125**, 156-165.

Mizrahi, V., Dawes, S.S. & Rubin, H. 2000. DNA Replication. In *Molecular Genetics of Mycobacteria.* (eds. G.F. Hatfull, and W.R. Jacobs, Jr.), pp. 159-172. ASM Press, Washington, DC.

Mukhapadhyay, D., Sen, U., Zapf, J. & Varughese, K.I. (2004). *Acta Crystallogr* **D60**, 638-645.

Murshudov, G.N., Vagin, A.A. & Dodson, E.D. (1997). *Acta Crystallogr* **D53**, 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 and Lung Disease* **79**, 3-29.

Reddy, V., Swanson, S., Sacchettini, J.C., Kantardjieff, K.A., Segelke, B. & Rupp, B. (2003). *Acta Crystallogr* **D59**, 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 Crystallogr* **D58**, 1514-1518.

Schapira, M., Abagyan, R. & Totrov, M. (2003a). *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. (2003b). *Proc Nat 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 Crystal Growth* **232**, 553-562.

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.

Tarantino, P.M., Zhi, C., Gambino, J.J., Wright, G.E. & Brown, N.C. (1999). *J Med Chem* **42**, 2035-2040.

Terwilliger, T.C., Park, M.S., Waldo, G.S., Berendzen, J., Hung, L.-W., Kim, C.-Y., Smith, C.V., Sacchettini, J.C., Bellinzoni, M., Bossi, R., et al. (2003). *Tuberculosis* **83**, 223-249.

Tomchick, D.R., Turner, R.J., Switzer, R.L. & Smith, J.L. (1998). *Structure* **6**, 337-350.

Turner, R.J., Bonner, E.R., Grabner, G.K. & Switzer, R.L. (1998). *J Biol Chem* **273**.

Van de Casteele, M., Chen, P., Roovers, M., Legrain, C. & Glansdorff, N. (1997). *J Bacteriol* **179**, 3470-3481.

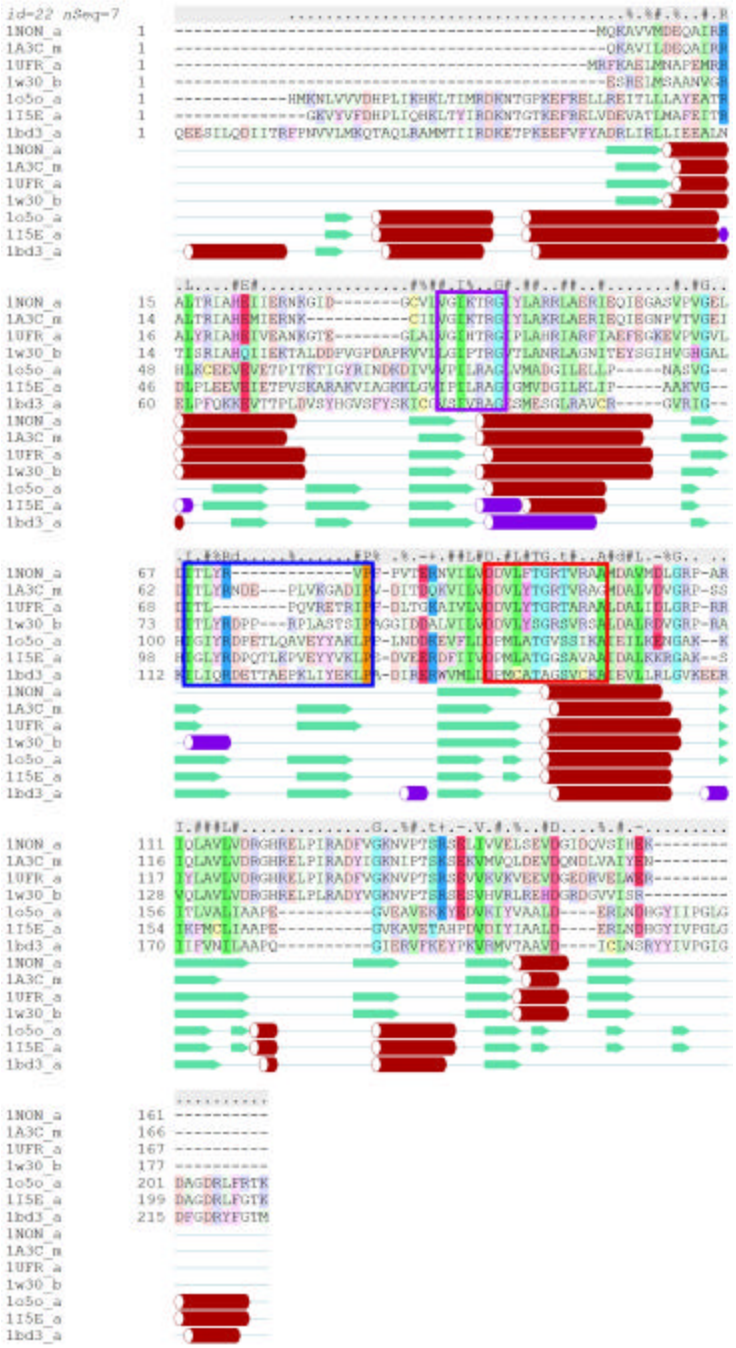

Figure 1 Multiple alignment showing sequence and structural homology between UPRTases and pyrRs, identified by PDB ID. Coloring is combination of amino acid type and consensus strength. Consensus sequence shown on top,. Helices are shown as red (alpha) or purple (310) cylinders, beta strands as green arrows. PRPP binding motif is boxed in red, Mg^{2+} loop boxed in purple, and flexible loop, which caps active site upon substrate binding, boxed in blue. Residue numbering is based on sequential extraction from PDB coordinates and may not be equivalent to numbering in full sequence. Alignments calculated and rendered wit h ICM-Pro 3.1.02 (Abagyan et al. 1994; Schapira et al. 2003a; Schapira et al. 2003b).

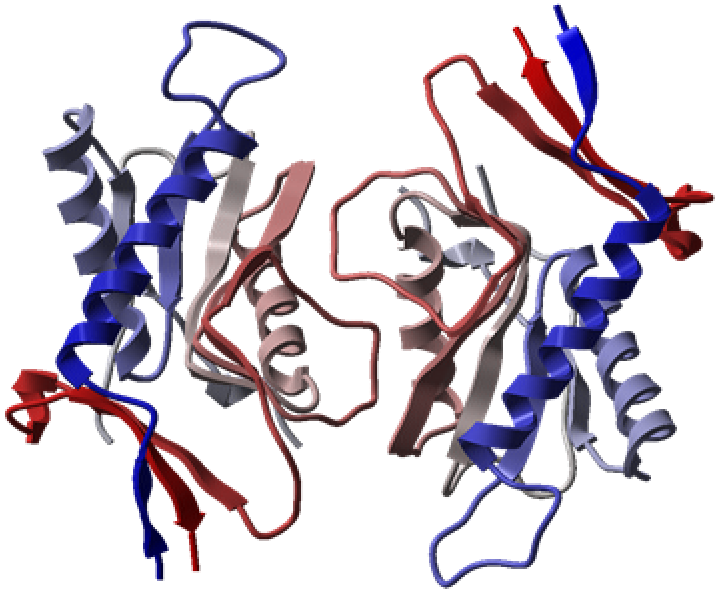

Figure 2 **Mtb pyrR homodimer.** Each monomer consists of a "core" parallel beta-sheet domain and a secondary "hood" domain. The core is reminiscent of the closely related UPRTases, whereas the hood is predominantly alpha helical in the UPRTases, the hood in pyrR terminates in a beta strand, forming an antiparallel beta sheet with the C-terminus. PyrR subunits also dimerize differently than the UPRTases. Coloring is blue to red, N-terminus to C-terminus. The buried surface area upon dimer formation is 750Å^2 per monomer. Rendered with ICM-Pro 3.1.02.

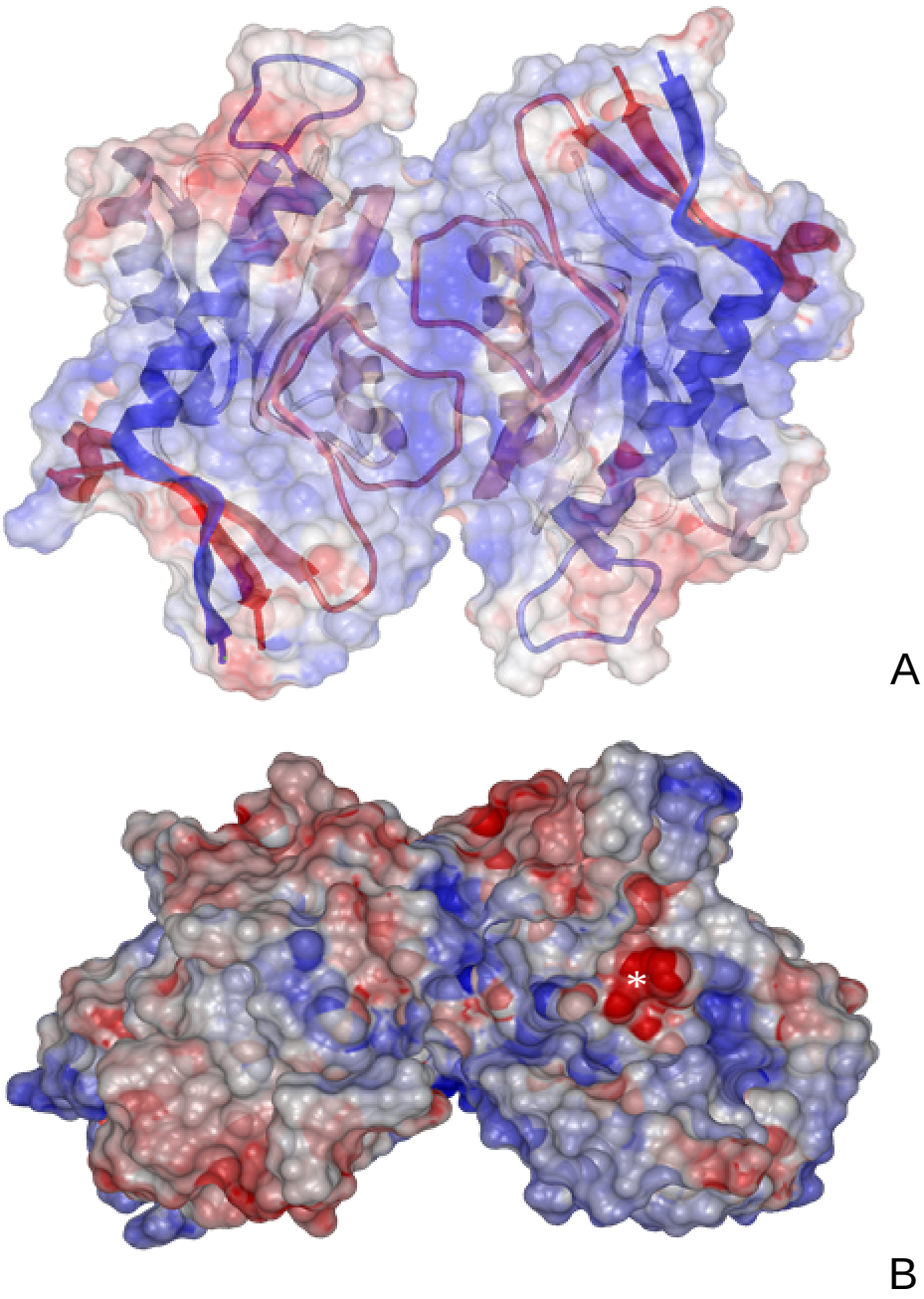

Figure 3 **Electrostatic potential surfaces.** (A) The electrostatic potential surface reveals that the Mtb pyrR homodimer possesses a deep positively charged cleft adapted to mRNA binding, as its primary function to regulate pyrimidine biosynthesis. (B) Orientation in (A) rotated 90° towards the viewer. The PRPP-UMP binding region (*) is easily located by structural homology to the active site of UPRTases.

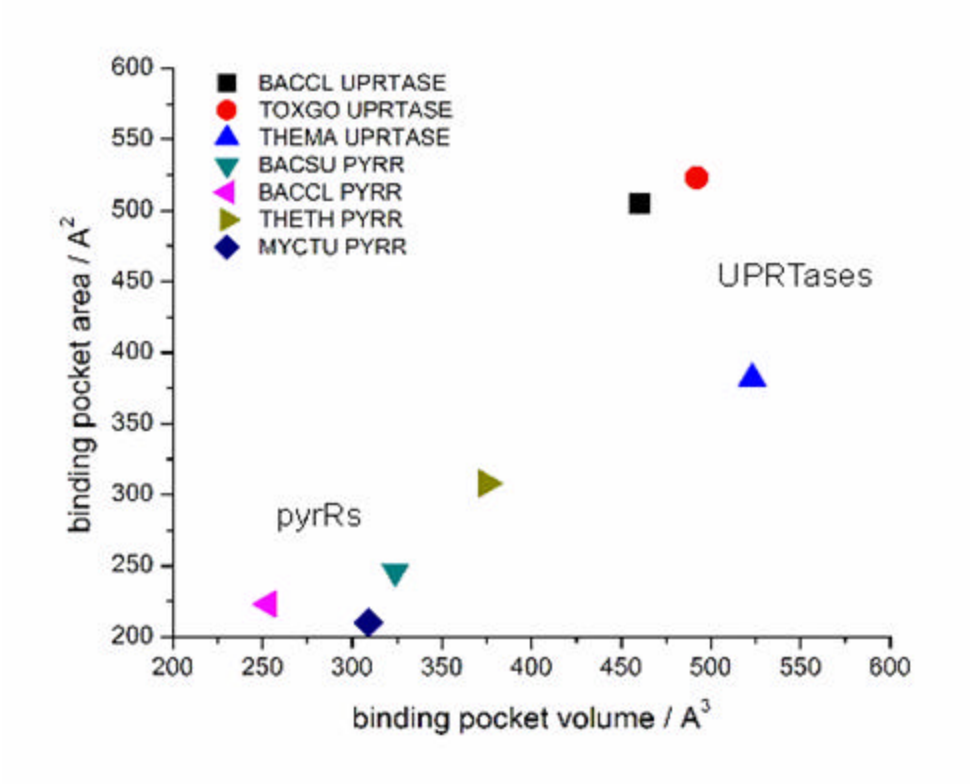

Figure 4 **Comparision of binding pocket characteristics in pyrRs and UPRTases**. A plot of surface area vs volume of binding pockets, as calculated by ICMPocketFinder (Abagyan et al. 1994; Schapira et al. 2003a; Schapira et al. 2003b), shows that pyrRs and UPRTases fall into two distinct groups. PyrRs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

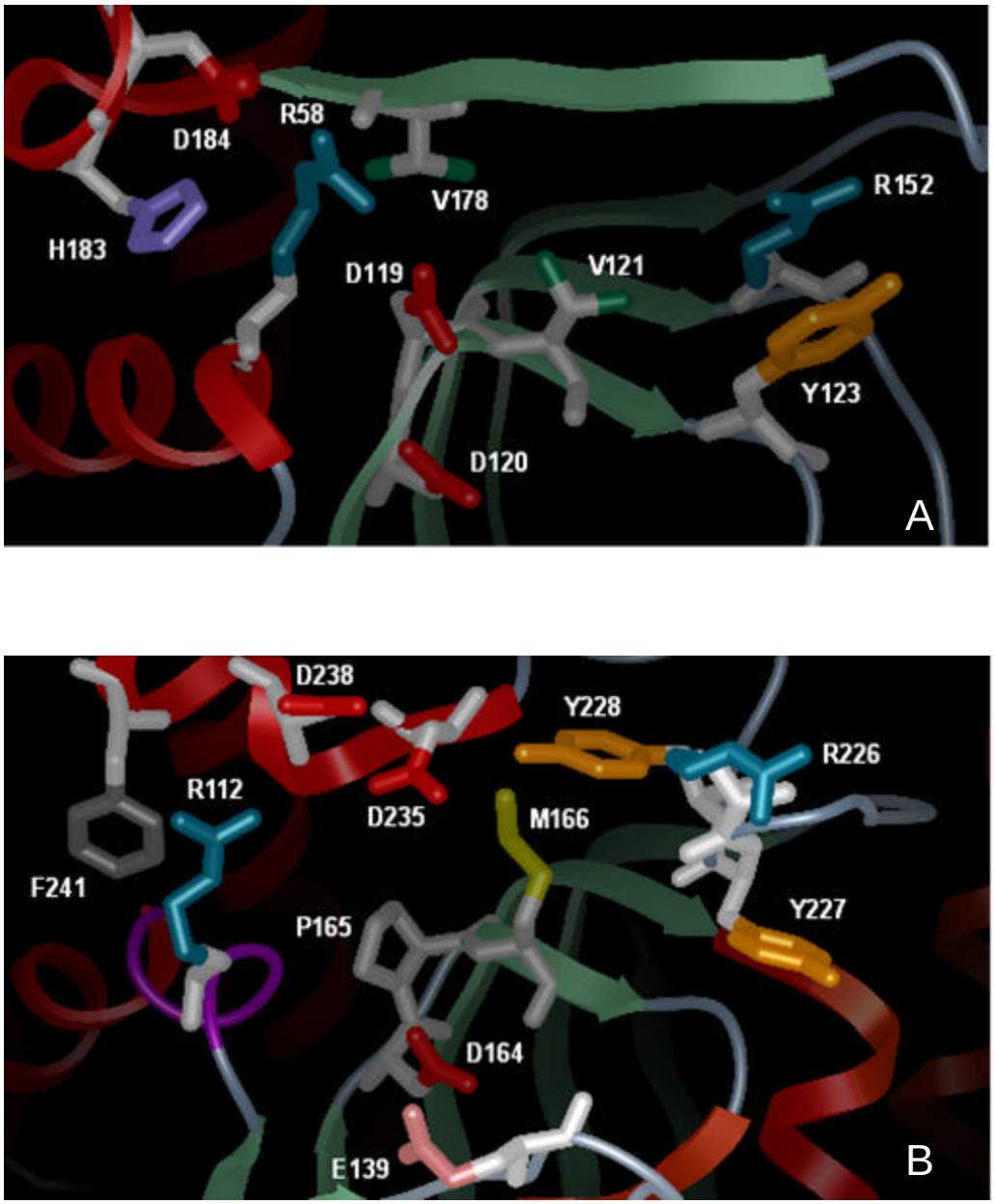

Figure 5 View into binding pockets of Mtb pyrR and UPRTase from *T. gondii*. (A) The binding pocket of Mtb PyrR contains conserved residues in pyrRs and UPRTases, including a conserved aspartate (119) involved in binding the ribosyl moiety of PRPP and UMP. C-terminal conserved aspartates (184, 187), proposed to stabilize the transition state (Kadziola et al. 2002), are not structurally conserved. Bulky hydrophobic residues in the hood, 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 hood domain, as well as aspartates that stabilize uracil binding and transition state (D164, D235 and D238) (Schumacher et al. 1998). Residues colored by type. Rendered with ICM-Pro 3.1.02 (Abagyan et al. 1994; Schapira et al. 2003a; Schapira et al. 2003b).

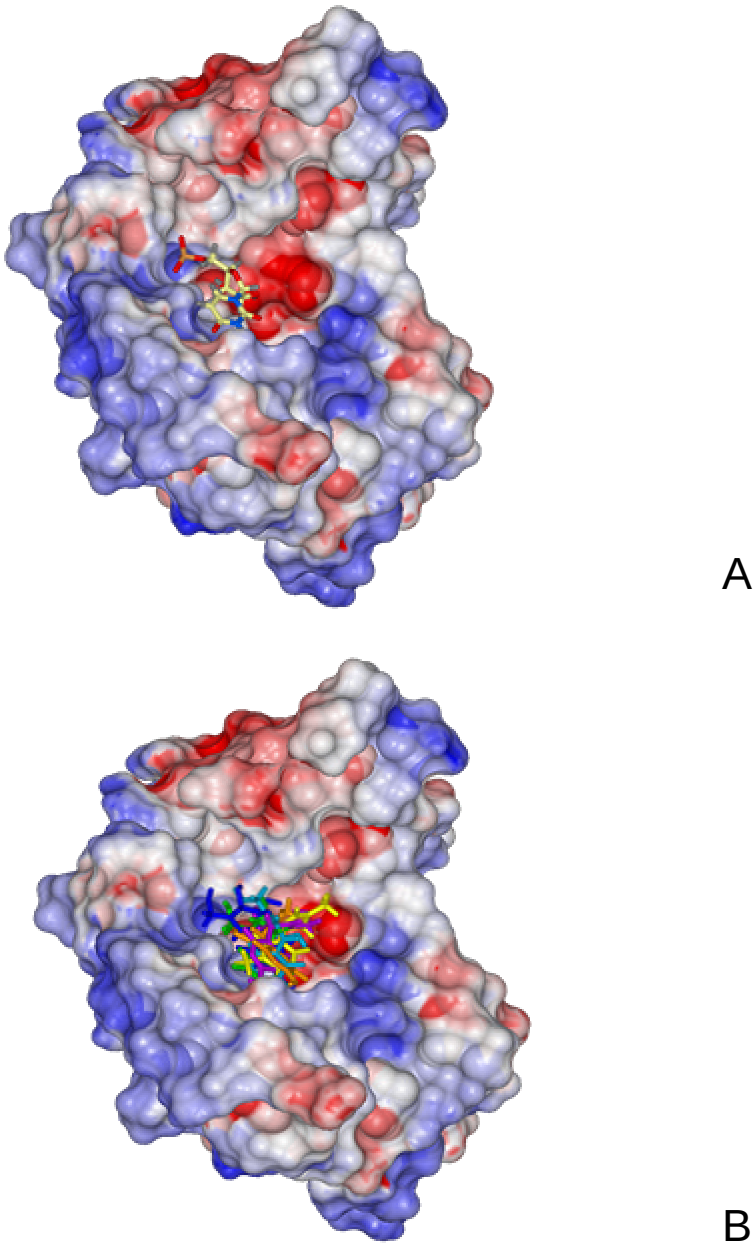

Figure 6 Virtual ligand screening against Mtb pyrR. (A) The predicted binding mode of UMP to Mtb pyrR is similar to that seen in the UPRTases. The binding pocket of Mtb PyrR contains many of the conserved residues in pyrRs and UPRTases, including a conserved aspartate (119), involved in binding the ribosyl moiety of PRPP and UMP. Orientation is same as Figure 2B. (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 3.1.02 (Abagyan et al. 1994; Schapira et al. 2003a; Schapira et al. 2003b). Shown are a series of high scoring L- and D- cyclopentenyl nucleosides. Efforts are underway 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.

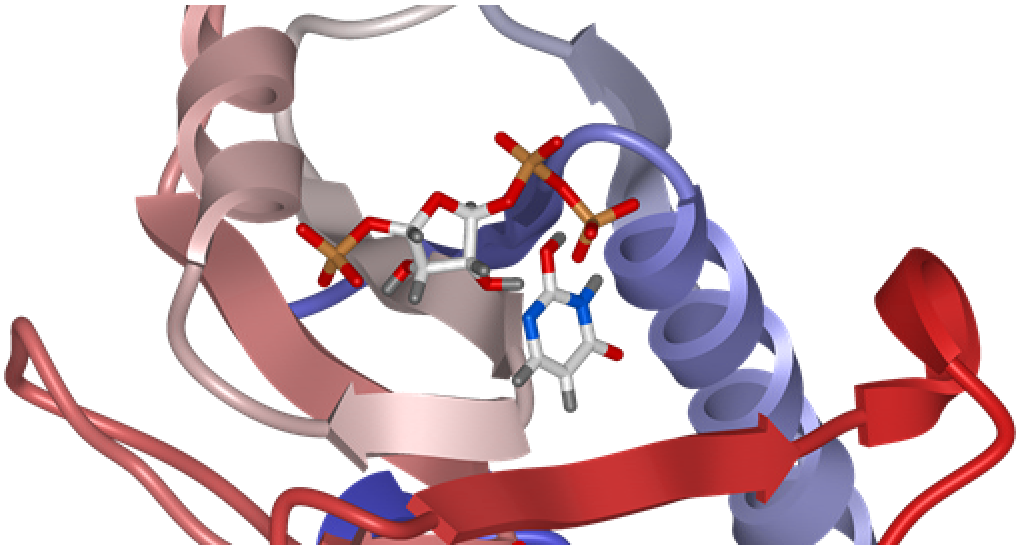

Figure 7 Simulated docking of PRPP and uracil to Mtb pyrR. The UPRTases have a conserved 13-residue fingerprint region, the PRPP binding motif, which is similar in pyrRs, and consists structurally of the beta strands and C-terminal helix, shown in pink (see also Figure 1). Sequential binding of PRPP and uracil to Mtb pyrR was simulated with ICM-Pro 3.1.02, and high-scoring conformations were refined by explicit global optimization of surface side-chains (Abagyan et al. 1994; Schapira et al. 2003a; Schapira et al. 2003b). 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 beta strand bottom of figure; see also Figure 5), which 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) (D184 and D187 located in the short helix and subsequent loop shown in red) are not structurally conserved. Coloring is blue to red, N- to C- terminus. Rendered with ICM-Pro 3.1.02 (Abagyan et al. 1994; Schapira et al. 2003a; Schapira et al. 2003b).

Table 1 Data collection and refinement statistics

Data Collection	
Space Group	$P3_221$
Wavelength (Å)	1.000
a,b (Å)	66.64
c (Å)	154.72
Resolution (Å)	38.8 – 1.85
Unique reflections[a]	34571 (3402)
Redundancy [a]	4.4 (4.0)
Completeness [a]	99.6 (99.8)
Rsym [a]	0.071 (0.423)
$< I/s_{(I)} >$ [a]	13.0 (2.4)
Number of molecules in a.u.	2
Vm (Matthews Coefficient)	2.3
% Solvent	46.5
Refinement	
Free R value [a], random, 5%	0.224 (0.290)
R value [a]	0.206 (0.278)
rmsd bond length (Å)[b]	0.020
rmsd bond angle (°) [b]	1.798
Overall coordinate error (Å) [c]	0.160
RSCC (*Shake&wARP*)[d]	0.90
RSCC (*Refmac5*)[e]	0.96
Ramachandran appearance [f]	
Most favored region (residues, %)	275 (93.9)
Additional allowed (residues, %)	15 (5.1)
Generously allowed residues, %)	3 (1.0)
Disallowed (residues, %)	0

[a] Values in parenthesis for the highest resolution bin (1.92 – 1.85Å)

[b] Deviations from restraint targets (Engh & Huber 1991)

[c] Estimated Standard Uncertainty, Diffraction Precision Index (DPI) based on R free (Cruickshank 1999)

[d] Real Space Correlation Coefficient, F_c map against averaged and weighted *Shake&wARP* map

[e] Real Space Correlation Coefficient, F_o map against F_c map, as reported by *Refmac5*

[f] Regions as defined in PROCHECK (Laskowski et al. 1993)