

Crystal structures of *Toxoplasma gondii* uracil phosphoribosyltransferase reveal the atomic basis of pyrimidine discrimination and prodrug binding

Maria A.Schumacher^{1,2}, Darrick Carter¹, Daniel M.Scott¹, David S.Roos³, Buddy Ullman¹ and Richard G.Brennan^{1,4}

¹Department of Biochemistry and Molecular Biology and ²Vollum Institute, Oregon Health Sciences University, Portland, OR 97201-3098 and ³Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

⁴Corresponding author
e-mail: brennanr@ohsu.edu

Uracil phosphoribosyltransferase (UPRTase) catalyzes the transfer of a ribosyl phosphate group from α -D-5-phosphoribosyl-1-pyrophosphate to the N1 nitrogen of uracil. The UPRTase from the opportunistic pathogen *Toxoplasma gondii* is a rational target for antiparasitic drug design. To aid in structure-based drug design studies against toxoplasmosis, the crystal structures of the *T.gondii* apo UPRTase (1.93 Å resolution), the UPRTase bound to its substrate, uracil (2.2 Å resolution), its product, UMP (2.5 Å resolution), and the prodrug, 5-fluorouracil (2.3 Å resolution), have been determined. These structures reveal that UPRTase recognizes uracil through polypeptide backbone hydrogen bonds to the uracil exocyclic O2 and endocyclic N3 atoms and a backbone–water–exocyclic O4 oxygen hydrogen bond. This stereochemical arrangement and the architecture of the uracil-binding pocket reveal why cytosine and pyrimidines with exocyclic substituents at ring position 5 larger than fluorine, including thymine, cannot bind to the enzyme. Strikingly, the *T.gondii* UPRTase contains a 22 residue insertion within the conserved PRTase fold that forms an extended anti-parallel β -arm. Leu92, at the tip of this arm, functions to cap the active site of its dimer mate, thereby inhibiting the escape of the substrate-binding water molecule.

Keywords: crystallography/structure-based drug design/*Toxoplasma gondii*/toxoplasmosis/uracil phosphoribosyltransferase

Introduction

The parasite *Toxoplasma gondii* is a ubiquitous protozoan which infects an estimated one-third of the US population and up to 90% of certain European populations. Risk factors for infection include preference for rare meat, contact with contaminated soil and exposure to infected cats. Although infection in healthy adults is usually asymptomatic, infection in neonates and immunocompromised individuals can lead to devastating consequences. Indeed, *T.gondii* is the causative agent of both toxoplasmic encephalitis, the most common cause of focal brain lesions in people with the acquired immunodeficiency

syndrome, and congenital toxoplasmosis, the leading cause of neurological birth defects in children. The current arsenal of drugs that has been used to treat and prophylax toxoplasmosis, although effective, is far from ideal, as these agents elicit numerous pernicious side effects (Brooks *et al.*, 1987; Luft and Remington, 1992). Thus, the need for novel and more efficacious therapeutic regimens is acute.

The pyrimidine and purine salvage pathways of protozoan pathogens, such as *T.gondii*, offer a multiplicity of potential therapeutic targets, as these pathways are highly disparate between parasite and host (Berens *et al.*, 1995) and can be manipulated to incorporate cytotoxic analogs selectively into the nucleotide pools of the parasite. Cytotoxic pyrimidine and purine analogs have been an extensively employed therapeutic tool in a variety of antineoplastic, antiviral, antiparasitic and immunosuppressive protocols. In *T.gondii*, the pyrimidine pathway is relatively uncomplicated. The parasite synthesizes pyrimidine nucleotides *de novo* (Schwartzmann and Pfefferkorn, 1977), and uracil phosphoribosyltransferase (UPRTase) is the only operative enzyme that salvages pre-formed pyrimidines to the nucleotide level (Pfefferkorn and Pfefferkorn, 1977; Iltzsch and Tankersley, 1994). Mutants deficient in UPRTase activity fail to incorporate radiolabeled uracil, uridine and deoxyuridine into nucleotides, whereas cytosine, cytidine, deoxycytidine and thymidine are not even incorporated by wild-type parasites (Pfefferkorn and Pfefferkorn, 1977). Because UPRTase is not expressed in mammalian cells, it represents an attractive target for rational drug design through the incorporation of prodrugs that are lethal specifically to the parasite. It has been shown that the *T.gondii* UPRTase binds a variety of pyrimidine analogs (Pfefferkorn *et al.*, 1989), and some have documented antitoxoplasmal activity (Pfefferkorn and Pfefferkorn, 1977; Natalini *et al.*, 1979; Iltzsch and Tankersley, 1994). Specifically, 5-fluorouracil is converted by the *T.gondii* UPRTase to the nucleotide level where it becomes toxic to the parasite, thus serving as a subversive substrate (Carter *et al.*, 1997). Unfortunately, 5-fluorouracil is also highly toxic to the human host, which limits its usefulness. Clearly, the design of new prodrugs specific to the *T.gondii* UPRTase is needed.

Toxoplasma gondii UPRTase (EC 2.4.2.9) is a 244-amino-acid protein (mol. wt 27 kDa) (Donald and Roos, 1995) that belongs to the family of phosphoribosyltransferases (PRTases) which carry out the biosynthesis of pyrimidine, pyridine and purine nucleotides as well as the synthesis of the amino acids tryptophan and histidine in lower eukaryotes and bacteria (Musick, 1981). These enzymes catalyze chemically similar reactions involving phosphoribosyl transfer from the substrate α -D-5-phosphoribosyl-1-pyrophosphate (PRPP). Specifically, *T.gondii* UPRTase catalyzes the transfer of a ribosyl phosphate

Results and discussion

Overall structure of *T.gondii* UPRTase

The crystal structure of *T.gondii* apo UPRTase was solved by the method of multiple isomorphous replacement (MIR). To overcome several experimental problems associated with the wild-type protein, a mutant of UPRTase, C128V, was created in which residue Cys128 was substituted with valine. This mutant protein is far less oxidation sensitive than the wild-type protein but exhibits kinetic properties indistinguishable from those of the wild-type enzyme. Furthermore, C128V and the wild-type protein crystallize isomorphously.

The crystal structure of the *T.gondii* UPRTase reveals the presence of four monomers that comprise two tight dimers. The first 20 residues are disordered, and the structure includes residues 21–244 of each subunit. The latter residues can be divided into two main regions; a 'core' region (A1–B1–A2–A3–B4–A4–B5–B6–B7–A5–B8–A6–B9; where A represents α -helices and B represents β -strands), which contains the conserved PRTase fold observed in other class I PRTases, and a 'hood' which includes residues critical for pyrimidine binding and specificity (B10–B11–A7) (Figure 2A). Intriguingly, in addition to these regions, the UPRTase structure contains an insertion within the canonical PRTase fold between A3 and B4 (B2–B3) that forms an antiparallel β -arm (residues 82–103) which extends radically from the core of the protein to interact with its dimer mate (Figure 2B). This results in unusual dimensions of $35 \times 43 \times 68$ Å for the monomer.

The active site of the enzyme, identified through substrate and product soaks into pre-formed crystals, is a cleft formed by the hood and the PRPP-binding motif, which in UPRTase is composed of the C-terminus of B7, the turn that follows and the N-terminus of A5. In the apoenzyme, a phosphate or sulfate molecule (both of which were used in crystallization) is bound to the N-terminus of A5 (Figure 2A). Structures of UPRTase obtained with ammonium sulfate contain an additional sulfate ion located near the N-terminus of A4.

Dimerization and the β -arm

The four monomers of the UPRTase in the asymmetric unit are nearly identical, with r.m.s.d.s (root mean square deviations) of 0.17, 0.40 and 0.38 Å for pairwise comparisons of the C_α atoms of residues 21–243. The tetramer displays 222 symmetry and forms a dimer of dimers. The slight differences between the r.m.s.d.s between dimer pairs can be attributed to differences in the region from B5 to B6, which constitutes the so-called flexible loop of class I PRTases (see below). Dynamic light scattering (DLS) experiments indicate that the protein is a dimer or salt-dependent tetramer in solution (data not shown). The dimer is oblong in shape with dimensions of $35 \times 43 \times 84$ Å. The dimer interface is bipartite and involves a long antiparallel β -arm, formed by B2 and B3, and residues from A2, A3 and the turn following α -helix, A4 (Figure 2B). The interface is extensive and buries 2010 Å² of each monomer surface from the solvent, 60% of which is hydrophobic. Although the dimer is essential for substrate binding (see below), the biological relevance of the tetramer is not yet clear.

Specific residues that contribute to the interface arise from the N-terminus of A3 of one subunit and residues from the C-terminal portion of the turn following A4' (where the prime indicates the other subunit in the molecular dimer) of the other subunit. Contacts are also provided by residues of A3 from one subunit to A3'. These contacts bury 745 Å² of each monomer from solvent. This part of the interface is highly hydrophilic and includes several tightly bound water molecules sandwiched between the subunits. Interestingly, there is an abundance of contacts involving arginine and acidic residues. Interactions in this region include those between the side chains of Arg126, from the turn following A4, to Glu61 located at the N-terminus of A3; stacking interactions between the side chains of Arg126 and Phe64', also located on A3; contacts between Glu75, from A3 to Arg68 on A3'; and stacking interactions between the side chains of Arg71 and Arg71'. Asn79, at the C-terminus of A3, hydrogen-bonds with the amide nitrogens of Ala43' and Gln44 from the N-terminus of A2, effectively capping the C-terminus of A3'.

Critical to the dimerization of UPRTase is its protruding β -arm, which cements the dimer by forming a tight interface with helix A2 from the other subunit (Figures 2B and 3). These contacts bury 1265 Å² of accessible surface area of each monomer. In contrast to the other contact surface, this interface is more hydrophobic. A key interaction of this extensive interface is the insertion of Tyr96 of the β -arm between the side chains of Arg46' and Arg53' of A2' (Figure 3). Phe83 and Phe101, at the top of the β -arm, stack against the N-terminus of A2', while Pro91 at the tip of the β -arm dips into the C-terminus of A2'. The side chain of Thr90, also near the tip of the arm, hydrogen-bonds with the carbonyl oxygens of Pro231' and Gly232' of the hood. These interactions allow the side chain of Leu92 to reach into the active site of its dimer mate and thereby block the bulk solvent access to the active site.

The dramatic extrusion of the β -arm, which protrudes over 35 Å from the protein core, clearly defines its role as an oligomerization domain. Although a highly unusual structure, β -arms have been observed in other proteins, including the *Bacillus stearothermophilus* histone-like protein, HU (Tanaka *et al.*, 1984; Vis *et al.*, 1995), the bacteriophage λ integration host factor, IHF (Rice *et al.*, 1996), the *E.coli* DNA-binding protein, Fis (Safo *et al.*, 1997), the ribosomal S7 protein (Hoaska *et al.*, 1997; Wimberly *et al.*, 1997) and chymotrypsin (Newman *et al.*, 1991). HU and IHF utilize their β -arms for DNA binding and, recently, IHF was seen to use a proline residue at the tip of the β -arm to contact its DNA-binding site specifically. The ribosomal S7 protein appears to use its β -arm similarly, but to contact RNA. On the other hand, the β -arm of chymotrypsin serves to cover its active site. UPRTase appears to be the first example in which a β -arm is utilized both as an active site cap and an oligomerization domain.

Comparison of the UPRTase core with other class I PRTases

Class I PRTases exhibit very little sequence homology, yet structural studies on members of this family have revealed a core that is common to all. Comparison of the structure of the *T.gondii* UPRTase with other class I

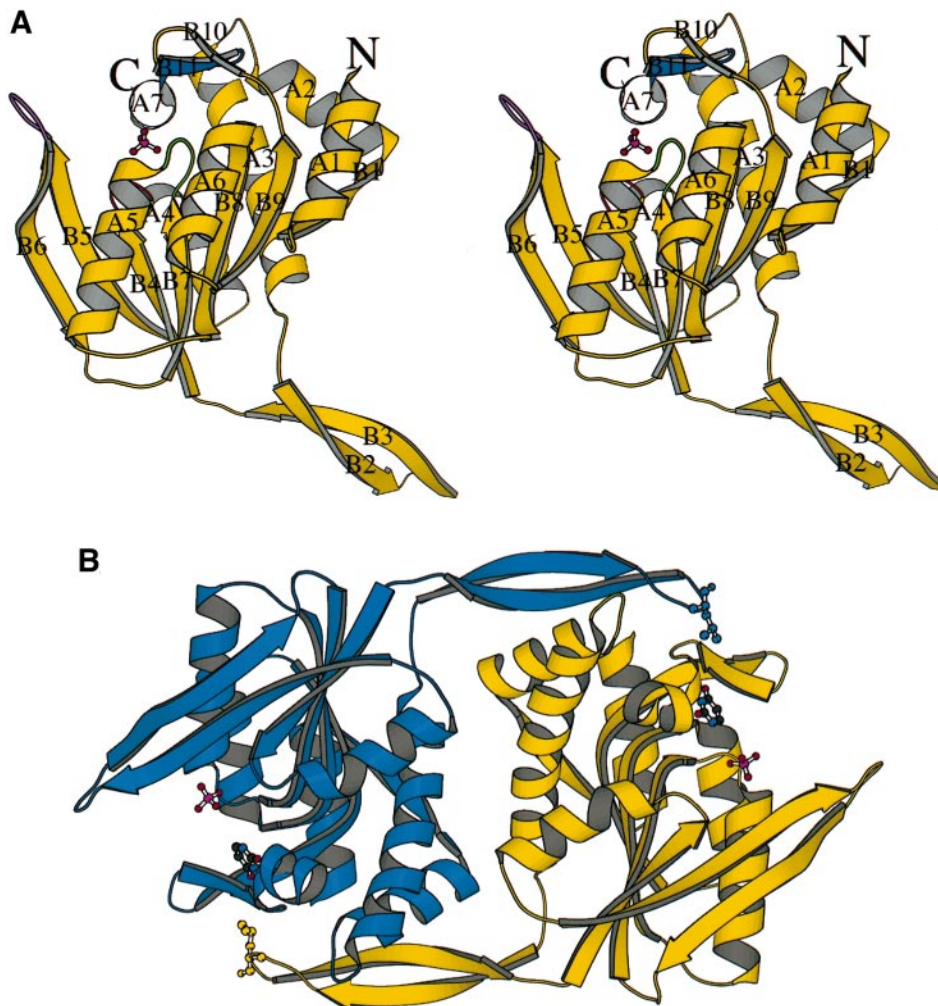


Fig. 2. (A) Stereo view of the *T.gondii* apo UPRTase monomer displayed as a ribbon diagram. Shown in ball-and-stick representation is a bound phosphate molecule in the catalytic pocket. The secondary structural elements are labeled using B for β -strands and A for α -helices. β -strands include B1 (36–42), B2 (82–89), B3 (94–102), B4 (104–111), B5 (128–138), B6 (143–152), B7 (158–167), B8 (188–198), B9 (212–219), B10 (222–224) and B11 (227–230). α -helices are A1 (21–33), A2 (42–53), A3 (58–78), A4 (114–124), A5 (170–183), A6 (198–209) and A7 (234–242). The regions of UPRTase that are the most highly conserved among all sequenced UPRTases are indicated in color: region I (*T.gondii* UPRTase residues 110–114) is shown in red; region II (residues 137–144) is shown in magenta; region III (residues 164–174) is shown in green; and region IV (residues 228–240) is shown in blue. Also labeled are the C-terminus (C) and the first observed N-terminal residue (N). The phosphate bound in the catalytic pocket is rendered as ball and stick (phosphorus, pink; oxygen, red). (B) Ribbon diagram of the *T.gondii* UPRTase dimer in which one monomer is colored yellow and the other blue. Represented in ball-and-stick (carbon, black; phosphorus, pink; nitrogen, blue; oxygen, red) are uracil and a phosphate ion, the latter of which is located identically to that seen in the apo UPRTase structure. Also shown in ball-and-stick (yellow and blue) is the Leu92 side chain which caps off the active site of its dimer mate. Both parts of this figure were generated with MOLSCRIPT (Kraulis, 1991).

PRTases identified four contiguous regions of similarity. These regions correspond to the previously identified PRTase core and consist of UPRTase residues 72–76, 105–113, 128–133, 158–169, 190–198 and 214–217 (coincident with UPRTase A3–B4–A4–B7–A5–B8–B9) with the most extensively conserved region corresponding to the PRPP-binding motif (UPRTase residues 158–169).

Superimposition of the 44 C_{α} atoms of these regions of UPRTase onto the analogous atoms of the other class I PRTases results in r.m.s.d.s of 1.44, 1.50, 1.27, 1.90, 1.32, 1.25 and 2.02 Å for the *E.coli* OPRTase, *S.typhimurium* OPRTase, human HGPRTase, *T.foetus* HGXPRTase, *T.gondii* HGXPRTase, *E.coli* XPRTase and *S.typhimurium* GAT, respectively. Structures of the HG(X)PRTases differ from other PRTases in having only three α -helices in their cores, while other PRTases, including UPRTase, have an

extra core α -helix (A6 in UPRTase) (Figure 2A). Perhaps the most striking difference between the structure of the *T.gondii* UPRTase core and those of the class I PRTases is the presence of the insertion of the 22-residue β -arm between α -helix A3 and β -strand B4. Also, unlike other class I PRTases, UPRTase does not contain a diacidic cluster in its PRPP-binding site but rather has a corresponding Asp–Pro dipeptide (residues 164 and 165). However, this substitution does not alter the structure of the region around this dipeptide, which in all class I PRTases is a turn.

Uracil binding and base substrate specificity

The *T.gondii* UPRTase binds uracil with high specificity, utilizing neither cytosine nor thymine as substrates. A systematic study, in which 100 compounds were evaluated as possible substrates for the *T.gondii* enzyme, underscored

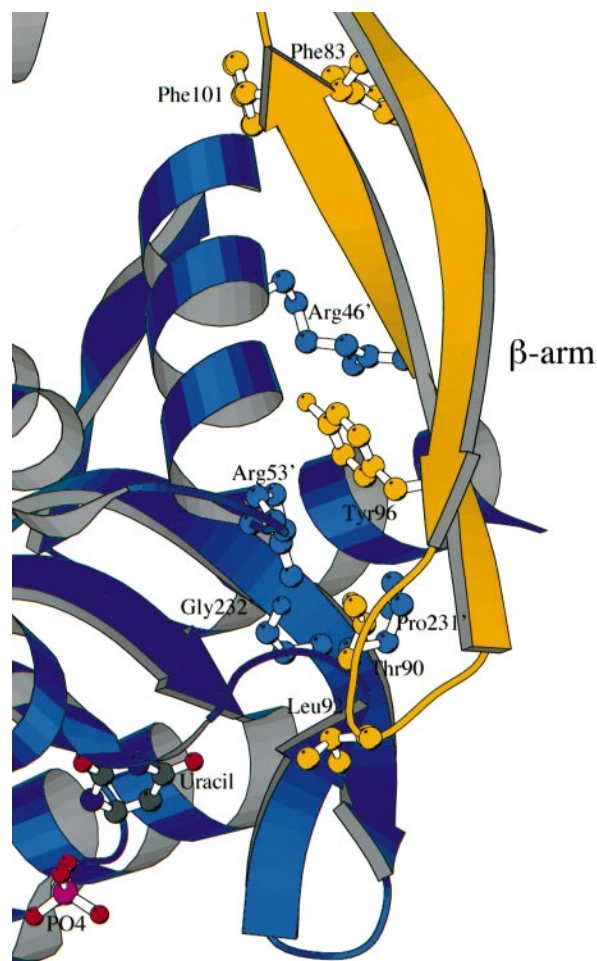


Fig. 3. Dimerization interface involving the β -arm. Shown are the contacts made by the β -arm (colored yellow) from one subunit to its dimer pair (colored blue). Only part of the β -arm is shown for clarity. Labeled are those residues making dimer contacts, and these include Phe83 and Phe101, which stack against the N-terminus of A2', Tyr96 which is sandwiched between Arg46' and Arg53', and Thr90 which makes hydrogen bonds to the carbonyls of Pro231' and Gly232'. Also shown is Leu92, which encloses the active site of the other monomer. Enzyme-bound uracil and phosphate are displayed and colored by atom type, whereby red is oxygen, blue is nitrogen, gray is carbon, and magenta is phosphorous.

the importance of the exocyclic O2, O4 and endocyclic N3 positions in binding (Iltzsch and Tankersley, 1994). That study also found that large exocyclic groups at position 5 of the pyrimidine base abolished binding. To elucidate the residues involved in substrate binding and specificity, uracil was soaked into crystals of the apoenzyme and the structure of the complex determined to 2.2 Å resolution.

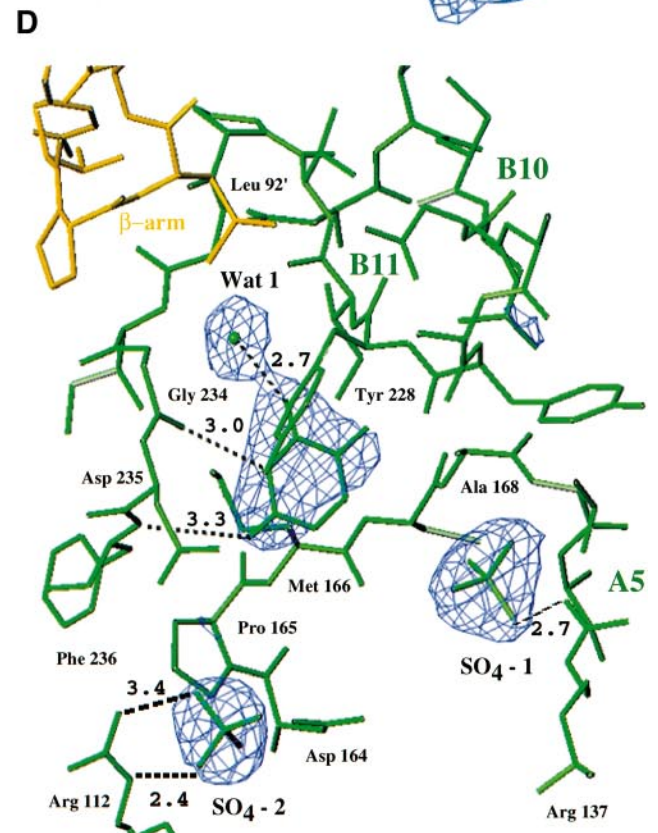
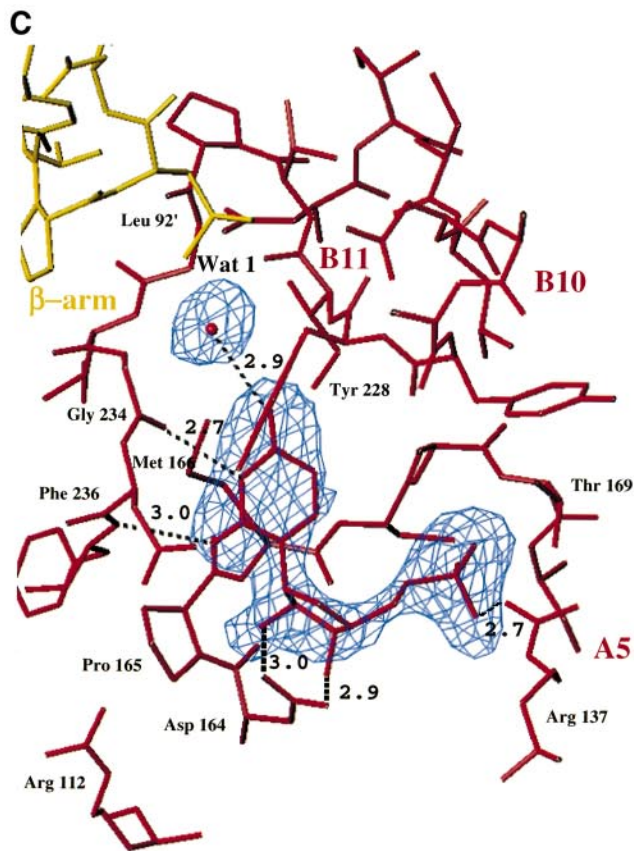
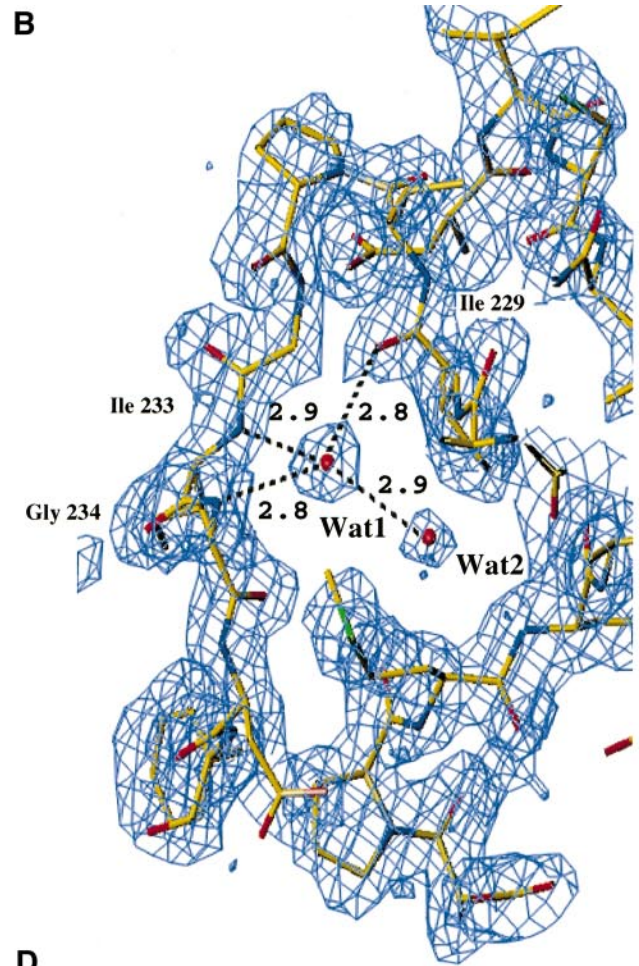
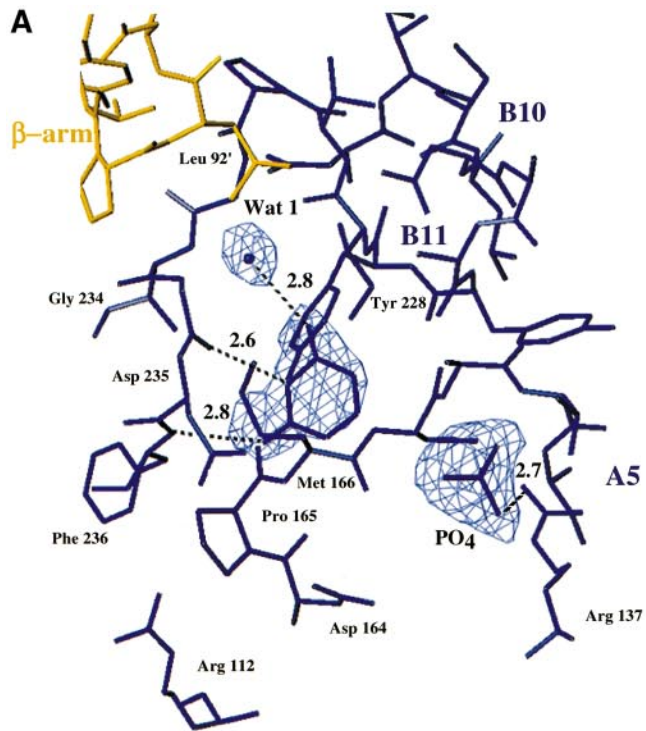
Superimposition of all C α carbons of each subunit of the uracil-bound UPRTase onto those of the apo structure results in an average r.m.s.d. of 0.20 Å, indicating that no large structural changes occur upon uracil binding. As observed in all four subunits, uracil is bound in the cleft formed by residues from the PRPP-binding motif in the core and residues from the hood (Figures 2B and 3). The exocyclic O2 and N3 ring atoms of uracil are anchored, β -strand fashion, against the backbone of residues Gly234 and Phe236 of the hood (Figure 4A). In this tight interaction, O2 hydrogen-bonds to the amide nitrogen of Phe236 (O2–NH, 2.84 Å) and N3 donates a hydrogen

bond to the carbonyl of Gly234 (CO–NH3, 2.62 Å). Unexpectedly, discrimination at the exocyclic O4 position involves a water-mediated contact in which a tightly bound water molecule (Wat1) donates a hydrogen bond to O4 (HOH–O4, 2.77 Å) (Figure 4A). This interaction is specific for uracil and discriminates against cytosine binding because the two hydrogen bond acceptors and one hydrogen bond donor of Wat1 are taken by the polypeptide backbone (Figure 4B); the acceptors by the amide nitrogens of Ile233 (H₂O–NH, 2.90 Å) and Gly234 (H₂O–NH, 2.83 Å) and the donor by the carbonyl oxygen of Ile229 (CO–HOH, 2.80 Å). This leaves only a single hydrogen bond donor to contact the exocyclic 4 position of the pyrimidine ring and limits the natural substrate to only uracil or thymine (Figure 4A). The exocyclic O4 oxygen of uracil also appears to make a weak hydrogen bond to the amide nitrogen of hood residue Ile229 (O4–NH, 3.12 Å) which may also aid in discriminating against 4-aminopyrimidines. The binding pocket around the endocyclic positions 5 and 6 is formed by residues Cys167 and Ala168 of the PRPP-binding motif as well as hood residues Tyr228 and Ile229. Thus, these residues as well as the rest of the uracil-binding residues effectively seal off most of the uracil-binding pocket from the bulk solvent.

It should be emphasized here that solvent molecule, Wat1, is not transiently associated with the protein and can be considered an extension of the protein. This water, which determines specificity at the exocyclic position 4, is present in the apo structures of all four monomers and displays low thermal parameters (average *B*, 29 Å²). In the apo UPRTase structure, in addition to Wat1, there are several water molecules within the substrate-binding pocket. Binding of uracil displaces one water molecule, Wat2, which is tightly bound to Wat1 (Wat1–Wat2, 2.9 Å) and thereby satisfies its second hydrogen bond donor in the absence of uracil (Figure 4B). Such water-mediated specificity is becoming increasingly appreciated in protein–ligand interaction (Frey, 1993; Baker, 1995). Highly analogous to the O4–water interaction described here for uracil–UPRTase binding is the water-mediated discrimination against purines with exocyclic O2 oxygens by the purine repressor (Schumacher *et al.*, 1997) and against alternative DNA-binding sites by the Trp repressor (Otwinski *et al.*, 1988; Joachimiak *et al.*, 1994).

UPRTase selectivity for its pyrimidine substrate is provided solely by backbone contacts and a water-mediated hydrogen bond. The backbone conformation of the protein is critical for its interactions with the pyrimidine N3 and O2 atoms and is sequence dependent, corresponding to the most conserved region amongst UPRTases (see below). Residues that are critical for the conformation of this region, and thus binding, are Pro231 and two conserved glycines, Gly232 and Gly234. Pro231 is in the *cis* peptide conformation, which is necessary for the tight turn from B11 into A7. In addition, because of their increased plasticity, the conserved glycines contribute to the unusual structure that is involved in binding the O2–N3–O4 edge of the pyrimidine ring. The specific binding conformation of Wat1, which reads the exocyclic O4 of uracil, is also dependent on the conformation of this turn.

In addition to substrate specificity contacts, favorable energetic contributions to base binding are provided by uracil–side chain stacking interactions with residues



Met166 from the PRPP-binding motif and Tyr228 from the hood (Figure 4A). Comparison of the apoenzyme with the uracil-bound enzyme reveals that upon uracil binding, the side chain of Met166 is displaced 0.7 Å, allowing the tight binding of the uracil base. Thus, the flexible and hydrophobic nature of the methionine side chain explains the conservation of this residue among UPRTases. A hydrogen bond from the Tyr228 phenolic side chain to the Oδ1 of Asp235, another conserved residue in the hood, not only positions the Tyr228 residue to stack over the uracil ring, but also acts to seal off this part of the binding site from solvent (Figures 4A and 5). This stabilizing hydrogen bond may clarify the high degree of conservation of a tyrosine residue at this position in UPRTases. However, the aromatic stacking role clearly takes precedence, as implied by the allowed substitution of phenylalanine for tyrosine in this position for the *Nicotiana tabacum* enzyme (see below; Figure 6).

An unexpected contributor to the catalytic site is the β-arm from the other subunit. The β-arm extends over the active site of its dimer pair to form a cap over residues 229–236. Although there are no direct contacts to the pyrimidine base from residues in this lid, it effectively shields the ‘top’ of the pyrimidine-binding site from the bulk solvent. Especially critical is Leu92', which sits directly above the base-specificity water molecule (Wat1), and thereby prevents its escape (Figure 4A).

Comparison of the UPRTase–uracil structure with the only other PRTase–pyrimidine complex (Ozturk *et al.*, 1995), the OPRTase–orotate structure, reveals that the pyrimidine bases are bound similarly, i.e. between the hood and the core of the protein. However, in the OPRTase–orotate structure, the base is shifted more deeply into the core of the protein between the β-strands corresponding to B7 and B8 in UPRTase. In contrast, uracil does not interact with residues from B8. This is probably due to the presence of the stacking residue Met166 which seals off any substrate access to B8. As observed in the structure of the UPRTase–uracil structure, most of the interactions between the orotate and OPRTase involve protein main chain atoms. However, in the latter complex, one specific contact is made between the side chain of Arg156 and the O4 oxygen of the orotate.

UMP binding

The structure of the UPRTase–UMP complex was obtained via soaking experiments. Superimposition of all C_α carbons of the UMP-bound form of UPRTase onto those of the native structure resulted in an average r.m.s.d. of 0.24 Å, indicating that there are no large structural differences between the apo and product-bound forms of the enzyme. The UMP product was identified readily in difference Fourier maps. Its uracil ring is *anti* and the ribose ring takes the C2' endo conformation (Figure 4C). The phosphate moiety of UMP interacts with the N-terminus of A5 of the PRPP-binding motif, with the positive helix dipole probably contributing to binding. The majority of hydrogen bonds made to the phosphate are from the backbone amide nitrogens of residues 168–172 of A5, with side chain interactions provided by the O_γ of Thr169 and the O_γ of Ser172. Two water molecules also bind to phosphate oxygens in all four subunits. Additionally, the side chain of Arg137 from the ‘flexible loop’ (located between B5 and B6) swings in to make a contact to a phosphate oxygen (Figure 4C). Interestingly, the same contacts to a bound phosphate (or sulfate group) are observed in the apo UPRTase, uracil–UPRTase and 5-fluorouracil–UPRTase structures (Figures 2A and B, and 4A and D). Upon binding, UMP displaces this group. The interaction between Arg137 and the phosphate (or sulfate) moiety may help stabilize the flexible loop, which is involved in crystal packing contacts, and thus explain the requirement for phosphate or sulfate in crystallization.

Unlike the uracil moiety, the ribose ring and phosphate group of the bound UMP are highly solvent accessible. Protein–ribose hydrogen bonds are formed between the O2' hydroxyl group of the ribose and the Oδ1 of the conserved residue Asp164 from the PRPP-binding motif (O2'–Oδ1, 3.0 Å). The ribose O3' hydroxyl group interacts with both the Oδ1 and Oδ2 atoms of Asp164 (O3'–Oδ1, 3.1 Å; O3'–Oδ2, 2.9 Å). Similar interactions between the ribosyl hydroxyls and the aspartic acid side chain corresponding to UPRTase residue Asp164 have been noted in the human HGPRTase–GMP complex (Scapin *et al.*, 1994) and the *T.foetus* HGXPRTase–GMP complex (Somoza *et al.*, 1996). In the human HGPRTase–GMP complex, an additional contact to a ribose hydroxyl is

Fig. 4. (A) An F_o-F_c ‘omit’ electron density map of the UPRTase–uracil complex (blue wire mesh) calculated with phases from the model in which Wat1, uracil and the phosphate ion (PO₄) were omitted, and followed by 20 cycles of *xyzb* refinement. The uracil and selected residues of the catalytic site of one monomer are represented as blue sticks and Wat1 as a blue sphere. Labeled are those residues that are involved in the binding of uracil to UPRTase, and these include Met166, Tyr228, Gly234 and Phe236. Key bond distances (Å) are given. Also labeled are the flexible loop residue Arg137, which contacts the phosphate ion, conserved region I residue Arg112, which assumes a *cis*-peptide bond, Pro165 from the PRPP-binding motif and Leu92' from the β-arm of the other subunit (yellow sticks), which seals off the ‘top’ of that active site. The location of nearby secondary structure elements B10, B11 and A5 are indicated as well. Key hydrogen bonds are drawn as thick dashed lines. The density map is contoured at 3.5 σ. (B) F_o-F_c omit electron density map of the apo UPRTase catalytic pocket (blue wire mesh) calculated with phases from the model in which, residues 164–170 and 226–230, and Wat1 and Wat2 were omitted, and followed by 20 cycles of *xyzb* refinement. Labeled are residues Ile228, Ile233 and Tyr234, and solvent molecules Wat1 and Wat2, the latter of which is necessary to maintain the tetrahedral coordination of conserved Wat1 in the apo structure. Bond distances (Å) to Wat1 are given. Each solvent molecule is shown as a red sphere, and the polypeptide chain atoms as red (oxygen), blue (nitrogen), yellow (carbon) and green (sulfur) sticks. Key hydrogen bonds are drawn as thick dashed lines. The density map is contoured at 4.0 σ. (C) F_o-F_c omit electron density map of the UPRTase–UMP complex (blue wire mesh) in which Wat1 and UMP were omitted, and followed by 20 cycles of *xyzb* refinement. UMP and residues of the catalytic pocket of one monomer are represented as red sticks. Those which interact with Wat1 (red sphere) or UMP are labeled, and bond distances (Å) given. Also labeled are residues Pro165 from the PRPP-binding loop, and Leu92' from the β-arm of the other dimer subunit (yellow sticks), which seals off the ‘top’ of the active site, and secondary structure elements B10, B11 and A5. Wat1 is shown as a red sphere. Hydrogen bonds are drawn as thick dashed lines. The density map is contoured at 3.5 σ. (D) F_o-F_c omit electron density map of the UPRTase–5-fluorouracil complex (blue wire mesh) in which Wat1, 5-fluorouracil and the two active site sulfate ions (SO₄-1 and SO₄-2) were omitted, and followed by 20 cycles of *xyzb* refinement. The 5-fluorouracil and residues of the catalytic pocket of one monomer are represented as green sticks. Those residues which interact with 5-fluorouracil, Wat1 (green sphere) and SO₄-2 (Arg112) are labeled, and bond distances (Å) are given. Also labeled are secondary structure elements B10, B11 and A5. Arg112 from conserved region I contacts SO₄-2. SO₄-1 is bound in the same location as PO₄ in the apo, uracil-bound and UMP-bound structures. Hydrogen bonds are drawn as thick dashed lines. The density map is contoured at 4.0 σ. All parts of this figure were generated using O (Jones *et al.*, 1991).

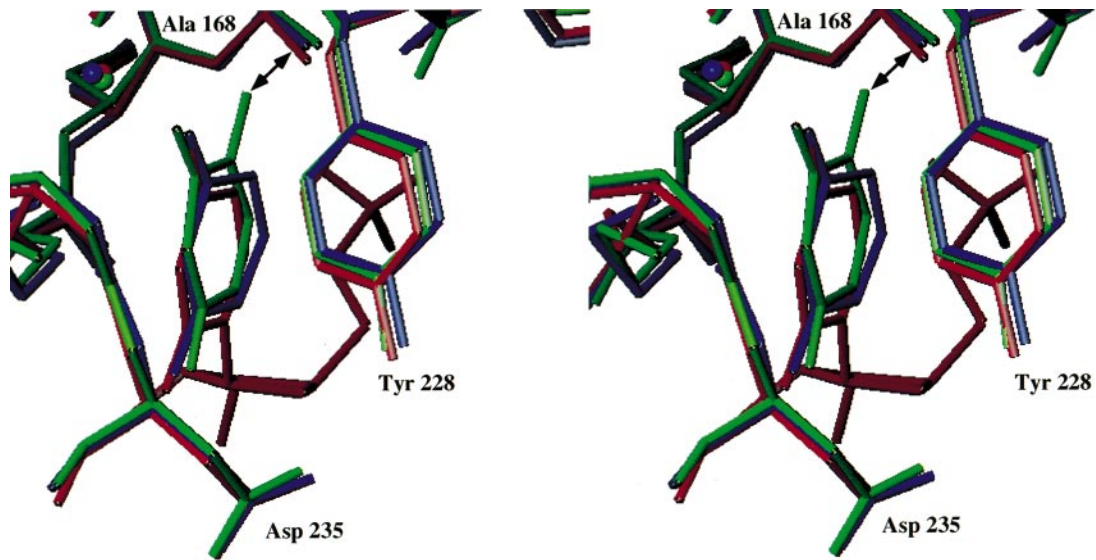


Fig. 5. Stereo view of the superimposition of UMP-bound UPRTase (red) and 5-fluorouracil-bound UPRTase (green) onto the uracil-bound UPRTase (blue). Note the rotation of the 5-fluorouracil ring that occurs to prevent steric clash between the fluorine atom and the C_β atom of Ala168. Also, the slightly ‘pulled out’ position of the uracil ring of the UMP moiety, compared with uracil, is revealed. Labeled are residues Ala168, Tyr228, which stacks over the uracil ring, and Asp235 which hydrogen bonds with the Tyr228. Wat1 of each complex is shown as an appropriately colored sphere. The double-headed arrow highlights the proximity of the fluorine atom of 5-fluorouracil to the C_β atom of Ala168. This figure was generated using O (Jones *et al.*, 1991).

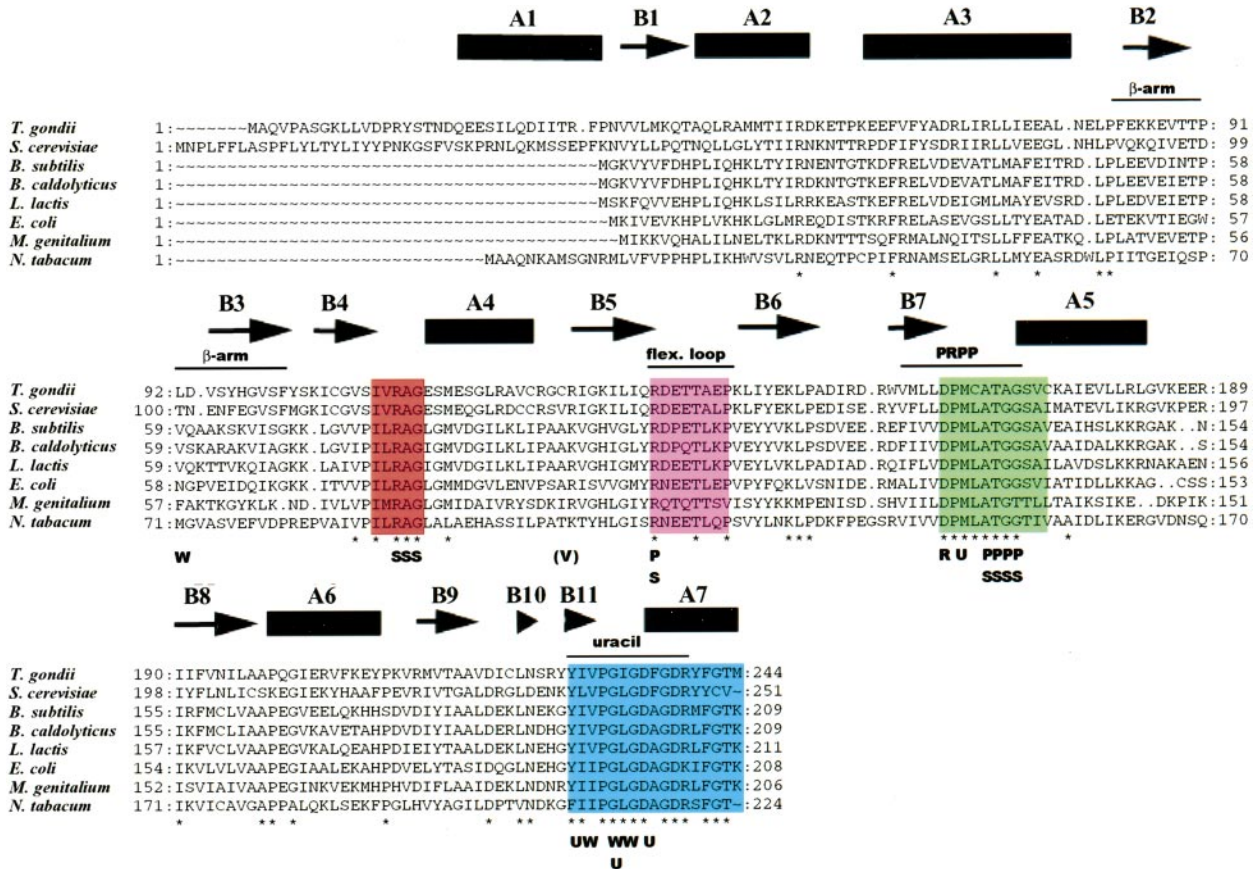


Fig. 6. Sequence alignment of representative UPRTases for which sequences are available. The sequence is colored to correspond to structure as depicted in Figure 2A. Conserved region I is shown in red, region II in magenta, region III in green and region IV in blue. The secondary structural elements of the *T.gondii* UPRTase are indicated above the sequence as black arrows for β-strands and black rectangles for α-helices, as are residues constituting the β-arm, flexible loop and PRPP, and uracil binding sites. Marked below are residue Cys128, which was mutated to a valine (V), the catalytic lid residue Leu92 (W), and residues involved in uracil binding (U), hydrogen bonding to Wat1 (W), phosphate (P) or sulfate binding (S), and ribose binding (R). A star indicates sequence identity in at least seven of the eight listed UPRTases.

made from the glutamic acid residue which follows the aspartic acid within the highly conserved diacidic cluster. The corresponding residue in UPRTase is Pro165.

As seen in the uracil-UPRTase complex, the uracil group of the bound UMP is involved in similar stacking interactions with Met166 and Tyr228 (Figure 4A and C). Moreover, the contacts between the exocyclic O2 and N3 atoms of the uracil moiety and the polypeptide backbone of residues Gly234 and Phe236 of the hood, and the discrimination at the exocyclic O4 position by Wat1 and the amide nitrogen of Ile229 are preserved (Gly234:CO-N3H, 2.7 Å; Phe236:NH-O2, 3.0 Å; HOH-O4, 2.9 Å; Ile229:NH-O4, 3.3 Å). Interestingly, superimposition of the substrate-bound and product-bound proteins reveals that the uracil moiety of the bound UMP is shifted relative to the bound uracil such that it is slightly more solvent exposed (Figures 4A and C, and 5).

Prodrug binding and discrimination against thymine: the UPRTase-5-fluorouracil complex

Previous studies have shown that pyrimidines containing substituents larger than fluorine at position 5, such as the methyl group of thymine, are not utilized as substrates by the *T.gondii* UPRTase (Iltzsch and Tankersley, 1994). However, the halogenated pyrimidine, 5-fluorouracil, is bound by UPRTase and converted to the nucleotide level by this enzyme. Thus, this pyrimidine functions as a subversive substrate. To provide a complete atomic description of the interactions between *T.gondii* UPRTase and 5-fluorouracil, as well as the stereochemical basis of its ability to discriminate against thymine, the structure of UPRTase in complex with 5-fluorouracil was determined (Figure 4D).

Structure determination of the UPRTase-5-fluorouracil complex revealed that the binding mode of this subversive substrate is nearly identical to that of uracil (Figure 4A and D). Binding of 5-fluorouracil caused no large structural changes in the protein, as indicated by the average r.m.s.d. of 0.21 Å between all C α s of all four subunits of the apo and 5-fluorouracil-bound proteins. As observed for uracil, the 5-fluorouracil is stacked between the side chains of Met166 and Tyr228 and makes the same specificity-determining hydrogen bonds to the polypeptide backbone and Wat1 (Gly234:CO-N3H, 3.0 Å; Phe236:NH-O2, 3.3 Å; HOH-O4, 2.7 Å; Ile229:NH-O4, 3.2 Å). However, in order to avoid steric clash between the 5-fluoro- group and the C β carbon atom of residue Ala168, the pyrimidine ring of 5-fluorouracil rotates (Figure 5). This rotation relocates the 5-fluoro group such that it is 3.0 Å from the C β of Ala168. Were the 5-fluorouracil to bind in the same manner as uracil, the distance between the fluorine atom and the C β of Ala168 would be prohibitively close (2.6–2.7 Å). The failure of 5-chloro-, 5-bromo- and 5-iodouracil to bind to the tightly constrained pyrimidine-binding pocket of UPRTase is the direct result of their longer bond lengths (C-Cl, 1.71 Å; C-Br, 1.87 Å; C-I, 2.07 Å), compared with 1.32 Å for C-F, and their much larger van der Waals radii. Modeling of these halogenated pyrimidines, in either the uracil- or 5-fluorouracil-binding conformation, shows steric clash between these and the C β of Ala168. Steric clash between this C β methyl group and that of thymine also prevents the binding of this natural pyrimidine. An alanine at this position is conserved in all UPRTases

and, correspondingly, all UPRTases studied thus far display the same discrimination against pyrimidines as that of the *T.gondii* enzyme. Further restriction is imposed by the carbonyl oxygen atom of hood residue Tyr228 (CO-F, 3.2 Å). Despite the ability of 5-fluorouracil to bind the *T.gondii* UPRTase, its affinity (as indicated by its relatively high K_i of 25 μ M, Carter *et al.*, 1997) is weaker than that of uracil and probably arises from the relatively fixed stereochemical constraints of the substrate-binding pocket (Figures 4D and 5).

Sequence conservation among UPRTases

Sequence alignments of multiple UPRTase protein sequences reveal a high degree of homology among these enzymes and clearly suggest that their three-dimensional structures will be very similar to the *T.gondii* enzyme (Figure 6). However, there are only four short regions that display sequence identity. Not surprisingly, these four regions, I–IV, encircle the active site of the *T.gondii* UPRTase enzyme and are key to substrate recognition and catalysis (Figures 2A and 6).

Region I (residues 110–114; red in Figures 2A and 6) forms the C-terminus of B4 and the turn that follows. Region II (residues 137–144; magenta in Figures 2A and 6) includes a loop between B5 and B6. Region III (residues 164–174; green in Figures 2A and 6) comprises the C-terminus of B7, a turn and the N-terminus of A5. Region IV (residues 228–240; blue in Figures 2A and 6) is composed of B11, a tight turn and A7. Sequences corresponding to region I in other class I PRTases have been implicated in PRPP binding, while region II contains the PRTase ‘flexible loop’ observed in all PRTase structures and shown recently to cover the active site (Smith *et al.*, 1997) and to participate in catalysis (Jardim and Ullman, 1997). Unlike most PRTase structures, the electron density for the loop is quite good and allows a complete trace of all loop residues in all four subunits. Region III contains residues of the PRPP-binding motif. Finally, region IV forms the part of the hood that is critical in pyrimidine binding.

In these four conserved regions, 13 of the completely conserved 16 residues are involved in substrate, product or phosphate/sulfate ion interactions, either through side chain or backbone contacts. Arg112, Ala113 (conserved region I), Arg137 (conserved region II), Ala168 and Thr169 (conserved region III) participate in phosphate (or sulfate) ion binding whilst, for UMP binding, residues Arg137, Ala168 and Thr169 as well as Asp164 (region III) are found to participate. Conserved region IV is critical for pyrimidine binding, with several residues including Gly234 contacting the base. Additionally, Met166 (region III) plays a crucial role in pyrimidine base stacking, as discussed above.

Residues that are identical among UPRTases that lie outside these conserved clusters are generally important for the fold or stability of the protein. Included are residues Phe62, Leu81 and Ile190, which are important in the formation of the core of the protein. Interestingly, conserved residue Arg53, from helix A2, reaches into the core to make a bifurcated hydrogen bond to conserved residue Asp220, which is located on the turn leading from the core into the hood. This interaction helps to anchor the hood against the core of the protein. In the *T.gondii*

enzyme, Arg53 also plays a critical role in dimer stability by stacking against Tyr96' of the β -hairpin from the other subunit (Figure 3). Conserved residues Leu72 and Glu75 also contribute to dimer stability by the stacking of the side chain of Leu72 against the hydrophobic part of the Glu75' side chain.

The sequence of the *T.gondii* UPRTase is the most similar to that of the *S.cerevisiae* enzyme (Natalini *et al.*, 1979). Both the *T.gondii* and *S.cerevisiae* enzymes have long N-terminal extensions of 37 and 41 residues, respectively. The *N.tabacum* UPRTase has a smaller 11 residue extension at its N-terminus (Figure 6). Interestingly, β -strand B1 of the *T.gondii* UPRTase, which corresponds to where all other UPRTase sequences begin to align, is the beginning of the core of the protein, indicating that the extra N-terminal residues in the yeast and *T.gondii* enzymes are not necessary for either the fold of the enzyme nor its enzymatic activity. Several of these 'extra' N-terminal residues (residues 2–20) are not observed in the electron density map for the *T.gondii* UPRTase, and the residues that are observed (21–32) form helix A1. This α -helix is stabilized through limited contact with A1' of a symmetry-related molecule, resulting in a crystallographic tetramer. Interestingly, while the *E.coli* enzyme is a homooligomer (Rasmussen *et al.*, 1986), the *S.cerevisiae* enzyme exists as an oligomer of two different subunits. Possibly, the 'extra' N-terminal residues in the *S.cerevisiae* and *T.gondii* enzymes play a role in hetero-oligomerization.

UPRTase active site architecture and catalytic mechanism

As has been observed in the structures of other class I PRTases, the *T.gondii* UPRTase active site is highly solvent exposed (Figure 2A and B). A conserved feature of all class I PRTases is a flexible loop which can move to cover the active site, thus shielding it from solvent (Krahn *et al.*, 1997). This movement is necessary for the formation of the oxocarbenium transition state, which otherwise would be highly unstable in a solvent-exposed environment. Several studies have implicated flexible loop residues in both PRPP binding and catalysis in class I PRTases. These include both mutagenesis and structural studies carried out on the *S.typhimurium* OPRTase and structural studies on *S.typhimurium* GAT (Krahn *et al.*, 1997) and *E.coli* OPRTase (Henriksen *et al.*, 1996). In the structure of GAT, loop residues Lys305 and Arg307 contact the bound AMP, while in *E.coli* OPRTase, loop residue Lys103 hydrogen-bonds to a sulfate bound at the active site. In *S.typhimurium* OPRTase, which is an obligate dimer, loop residues from one subunit contribute to the active site of the other subunit. Further support for both a catalytic and PRPP-binding role for flexible loop residues comes from studies on *E.coli* XPRTase which show that loss of an 11-residue peptide from its flexible loop by proteolysis resulted in a loss of catalytic activity, although the protein could still bind the product, GMP (Vos *et al.*, 1997). Subsequently, it was found that addition of PRPP to the protein significantly decreased protease digestion of the loop, consistent with a role for these residues in PRPP binding. Finally, recent studies have shown that mutation of flexible loop residues Ser95 and Tyr96 of the *Leishmania donovani* HGPRTase, which are conserved in all HG(X)PRTases, resulted in a two to three orders of magnitude loss in catalytic ability

(Jardim and Ullman, 1997). In *T.gondii* UPRTase, the flexible loop is positioned near the active site (Figure 2A and B). Two notable residues of the loop, Arg137 and Thr141, are conserved among UPRTases, which suggests their importance in either catalysis or substrate binding. The side chain of Arg137 contacts a bound phosphate in the apo and uracil complexes, a sulfate in the 5-fluorouracil complex and the phosphate group of UMP in the UPRTase–UMP structure (Figure 4A, C and D). Thr141 is located at the tip of the loop. However, repositioning of the loop over the active site would allow this residue to enter the catalytic pocket and thus play a role in catalysis.

Additional residues implicated in PRPP binding in class I PRTases are found in a tight turn which corresponds to residues 110–113 in *T.gondii* UPRTase. These residues are contained within conserved region I in UPRTases (Figures 2A and 6). Arg112 is particularly intriguing and takes a *cis*-peptide conformation. A *cis*-peptide conformation has been observed in the corresponding position in structures of the *T.foetus* HGXPRTase, *E.coli* OPRTase and *E.coli* XPRTase, and enables the formation of a hairpin turn between a β -strand and an α -helix (B4 and A4 in UPRTase). The significance of *cis*-peptide residues, specifically those of the type that do not involve prolines, in the catalytic centers and binding sites of proteins has been increasingly noted (Stewart *et al.*, 1990; Weiss *et al.*, 1998). Such strained peptide linkages have been proposed to represent a well of stored potential energy that may be important for catalysis. Interestingly, *E.coli* OPRTase and *E.coli* XPRTase contain bound sulfates near the *cis*-peptide residue. Similarly in structures of UPRTase crystallized in ammonium sulfate, in addition to the sulfate ion located at the N-terminus of A5, a second sulfate ion is found bound to the Arg112 side chain (Figure 4D). This sulfate interacts with amide nitrogens of the helix A4 and its helix dipole again probably contributing to binding. This anion binding by Arg112 suggests its potential involvement in PRPP binding. Further support for this proposal comes from X-ray data collected on UPRTase crystals which were soaked with PRPP (data not shown). The resulting $F_o - F_c$ electron density map revealed new, strong density only near Arg112. Unfortunately, the density was smeared and a PRPP molecule could not be refined in this location. The incomplete or improper binding to PRPP could be explained by the inability of the flexible loop residues to assume a conformation necessary for PRPP binding because of crystal packing.

The basis of pyrimidine base recognition by *T.gondii* UPRTase was revealed by structures of the UPRTase–uracil and UPRTase–5-fluorouracil complexes (Figure 4A and D). The ability to soak uracil and PRPP into crystals of apo UPRTase is consistent with the random order mechanism of substrate binding that was reported for *A.laidlawii* UPRTase (McIvor *et al.*, 1983). Examination of the apo UPRTase and UPRTase–UMP structures has identified three conserved acidic residues, Asp164, Asp235 and Asp238, near the active site that might be involved in formation or stabilization of the oxocarbenium transition state. Residue Asp238 forms an interaction with conserved residue Arg112. A display of the electrostatic potential surface of UPRTase reveals that these acidic residues combine to form a large negative patch within the active site, again implying their importance in catalysis

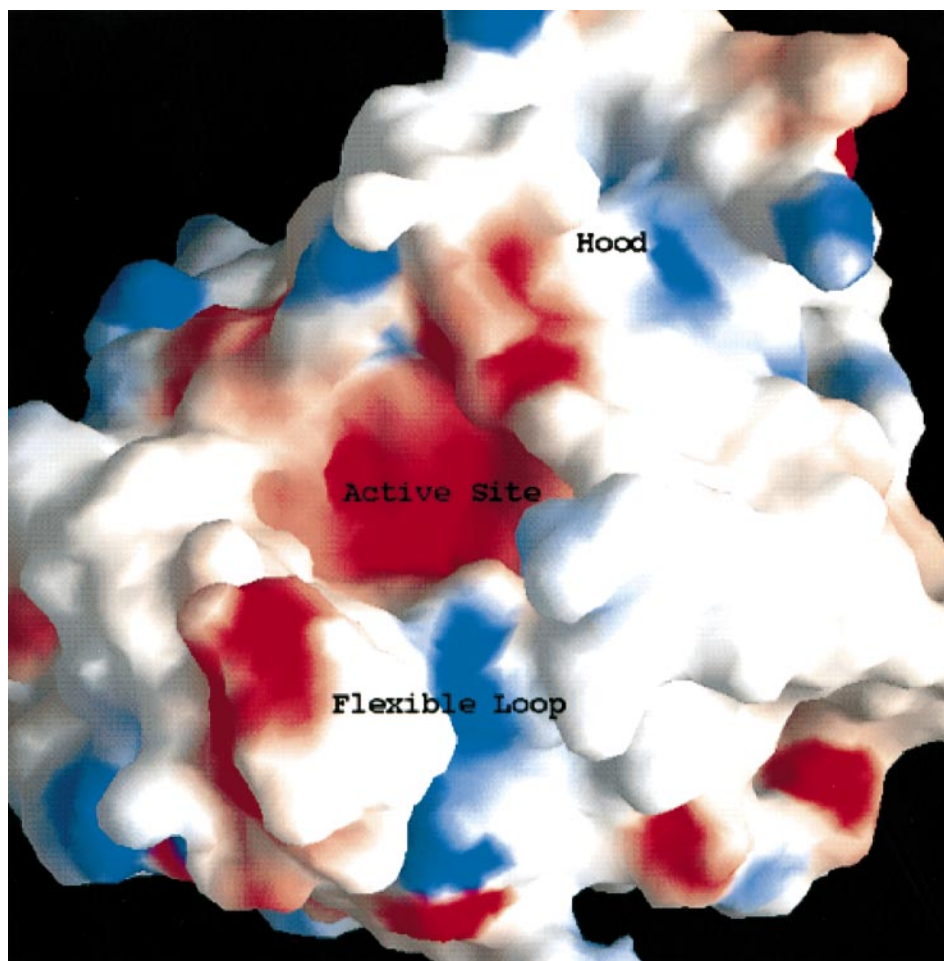


Fig. 7. Electrostatic potential surface in the region of the UPRTase active site. Red and blue regions represent negative and positive potentials, respectively. The active site, the flexible loop and the hood are labeled for reference. The large area of the negative potential in the active site is clearly visible and may serve to stabilize an oxocarbenium transition state. The areas of positive potential surrounding the negative potential may contribute to binding the negatively charged substrate, PRPP. The pyrimidine base-binding site is located under the hood. This figure was generated with GRASP (Nicholls *et al.*, 1991).

(Figure 7). Moreover, this negative patch suggests a mechanism for oxocarbenium ion formation in which binding of the negatively charged PRPP increases the pK_a value of one of the acidic residues and thereby allows the more facile abstraction of the proton from C1' (Figure 1). Two candidate residues are Asp164 and Asp235 as they are within 5 Å of the C1' atom of the ribose ring. The UPRTase surface also reveals a patch of positive charge around the active site that is contributed by residues Arg112 and Arg137 and could aid in the binding of the negatively charged substrate, PRPP. Clearly, the combination of molecular biological, kinetic and structural studies will be necessary to test the roles of these residues in the catalytic mechanism of UPRTase.

Conclusion

Current drugs and prophylaxis in the treatment of toxoplasmosis are far from ideal. The fact that humans do not possess a UPRTase activity coupled with the finding that the *T.gondii* UPRTase binds a variety of pyrimidine analogs, such as 5-fluorouracil, which have documented antitoxoplasmal activity, bolster the premise that the *T.gondii* UPRTase is a very good target for structure-based prodrug and drug design. The high-resolution structures

of the *T.gondii* apo UPRTase, UPRTase–uracil, UPRTase–UMP and UPRTase–5-fluorouracil complexes provide structural scaffolds with which to facilitate this endeavor. Moreover, these structures revealed the stereochemical bases of nucleobase binding and discrimination, which involves exclusively hydrogen bond donors and acceptors of the polypeptide backbone, a tightly bound water molecule and a constricted pyrimidine-binding pocket. Finally, because of the high sequence homology among all UPRTases, the *T.gondii* UPRTase structures described here should serve as models for the other UPRTases.

Materials and methods

Crystallization

A single mutation, C128V, was introduced into the *uprt* gene by standard methods, which changed residue Cys128 to a valine on the basis of the sequence alignment between the *T.gondii* and *S.cerevisiae* enzymes. This change mitigated multiple problems, in particular that of rapid and uncontrollable protein oxidation. Without this change, crystals were seen to crack over a period of a few days and change their unit cell dimensions during X-ray intensity data collection. The native and mutant proteins were purified using identical protocols (Carter *et al.*, 1997), displayed indistinguishable kinetic parameters (data not shown) and crystallized isomorphously. Further, structure determination revealed that residue 128 is present in a solvent-exposed turn far from the active site. Thus

Table I. Heavy atom data collection and statistics

	Native1	Thimerosal	Selenomethionine	Baker's dimercurial
Resolution (Å)	2.50	2.50	3.50	3.00
No. of measurements	91 555	83 961	42 852	57 604
No. of reflections	35 665	29 102	13 747	20 567
$R_{\text{sym}}(\%)^a$	5.4	3.9	7.8	10.4
$R_{\text{iso}}(\%)^b$		11.6	13.4	9.0
No. of sites		4	28	4
Phasing power ^c		1.72	1.08	1.13
R_c^d		0.655	0.689	0.676
Mean overall figure of merit ^e to 3.0 Å	0.50			

^a $R_{\text{sym}} = \Sigma|I_o - \langle I \rangle| / I_o$, where I_o = observed intensity and $\langle I \rangle$ = average intensity obtained from multiple observations of symmetry related reflections.

^b $R_{\text{iso}} = \Sigma||F_{\text{PH}}| - |F_{\text{P}}|| / \Sigma|F_{\text{P}}|$, where $|F_{\text{P}}|$ is the protein structure factor amplitude and $|F_{\text{PH}}|$ is the heavy atom derivative structure factor amplitude.

^cPhasing power = r.m.s. ($|F_{\text{H}}|/E$), where $|F_{\text{H}}|$ is the heavy atom structure factor amplitude and E is the residual lack of closure.

^d $R_c = \Sigma||F_{\text{der}} \pm F_{\text{nat}}| - |F_{\text{H(calc)}}|| / \Sigma|F_{\text{der}} - F_{\text{nat}}|$ for centric reflections, where $F_{\text{H(calc)}}$ is the calculated heavy atom structure factor.

^eFigure of merit = $\int P(\theta)\exp(i\theta)d\theta / \int P(\theta)d\theta$, where P is the probability distribution of θ , the phase angle.

we conclude that the structures of the *T.gondii* UPRTase and the C128V mutant are identical.

Crystals of the apo UPRTase were grown at room temperature via hanging-drop, vapor diffusion. UPRTase [in 10 mM Tris–HCl pH 7.6, 10 mM dithiothreitol (DTT)] was concentrated to between 25 and 40 mg/ml and the solution mixed with the reservoir solution, which was either 1.0 M ammonium phosphate, 0.1 M citrate pH 5.6, 0.2 M NaCl, or 1.0 M ammonium sulfate, 0.1 M Tris–HCl pH 7.6. Crystals obtained with these precipitants are isomorphous and grow within a period of 2 days to 1 month to dimensions as large as $1.0 \times 0.4 \times 0.4$ mm. Such crystals diffract in-house to 1.75 Å. The crystals were characterized by precession photography as monoclinic, space group $P2_1$ with unit cell parameters; $a = 60.6$ Å, $b = 141.8$ Å, $c = 71.4$ Å, $\beta = 115.0^\circ$. It was found that the monoclinic space group could be transformed to orthorhombic, space group $C222_1$, with $a = 60.6$ Å, $b = 129.5$ Å, $c = 141.8$ Å. For technical reasons, the crystal structure was solved in the monoclinic setting. The asymmetric unit of the monoclinic crystal contains four UPRTase molecules giving a V_m of $2.5 \text{ \AA}^3/\text{Da}$ and a solvent content of ~49% (Matthews, 1968).

Dynamic light scattering (DLS)

The oligomerization state of purified *T.gondii* C128V UPRTase was examined by DLS, at a concentration of 15 mg/ml, in 25 mM Tris–HCl, pH 7.6, 10 mM DTT, using a 2001 DynaPro Dynamic Light Scattering Instrument. Fifty measurements were taken at 25°C and analyzed by the DYNAMICS software, version 3.30. Bimodal analysis resulted in a molecular weight of 66 kDa, consistent with an elongated dimer. However, in the presence of ≥ 50 mM NaCl the protein is tetrameric.

Data collection

X-ray intensity data were collected at room temperature with an Area Detector Systems Corporation (ADSC) multiwire area detector (Xuong *et al.*, 1985) using a RIGAKU RU200-H rotating anode X-ray generator with graphite monochromator operating at 40 kV and 150 mA, or an R-AXIS IV imaging plate system using a RIGAKU RU300 rotating anode X-ray generator equipped with double focusing mirrors and operating at 50 kV and 100 mA (Tables I and II). Data collected on the system were processed with the software provided by ADSC, and data collected on the R-AXIS IV were processed using BIOTEX (Molecular Structure Corporation, Inc., Woodlands, TX).

Structure determination of apo UPRTase

The structure of the *T.gondii* apo UPRTase was solved by MIR using crystals soaked for 1 month in thimerosal (1mg/ml) or 2 weeks in Baker's dimercurial (0.5 mg/ml). A selenomethionine-substituted UPRTase was used as well and prepared as follows: an overnight culture of DL41 cells, transformed with the *T.gondii uprt-pBAce* construct, was resuspended in methionine-deficient, low phosphate induction medium supplemented with thymine (0.15 g/l), uracil (0.5 g/l), adenine (0.5 g/l) and guanosine (0.5 g/l). Prior to inoculation, 40 mg/l seleno-D,L-methionine was added to the medium. The DL41 cells were then grown in the dark for 48 h at 37°C and harvested by centrifugation. The selenomethionine UPRTase was purified as described (Carter *et al.*, 1997).

The initial difference Patterson maps were very difficult to interpret because of the presence of a strong peak, which arose from the packing

of the UPRTase molecules along a crystallographic axis. This large peak was observed in all difference Patterson maps. Eventually, four heavy atom sites were identified in the difference Patterson map of a thimerosal derivative. Assignment of one peak allowed the use of difference Fouriers to place the remaining three thimerosal peaks as well as the 28 selenomethionine peaks. The Baker's dimercurial-soaked crystal was derivatized at the same location as the thimerosal (Cys125). However, the occupancy of the sites was lower. The heavy atom parameters were refined via maximum likelihood using PHASES-95 (Furey and Swaminathan, 1997), resulting in an overall figure of merit of 0.50–3.0 Å (Table I). The initial MIR map was not interpretable, and solvent flattening, as implemented in PHASES-95, was used. The initial solvent-leveled map was interpretable and revealed the presence of a left-handed α -helix, at which point all coordinates were inverted and a new map calculated. Five β -strands and five α -helices were then evident for each subunit, and a polyalanine trace was fit. Iterative rounds of model building and phase combination resulted in a model that included residues 21–243, with an initial R -factor of 46.9% using data from 10 to 3.0 Å.

Model refinement: apo UPRTase, UPRTase–uracil, UPRTase–UMP and UPRTase–5-fluorouracil

Iterative cycles of positional least squares refinement with TNT (Tronrud *et al.*, 1985) and model rebuilding using FRODO (Jones, 1985) reduced the R -factor to 23.4% in the resolution range of 10.0–3.0 Å. At this point, tightly restrained B -factor refinement was begun and the resolution was extended slowly to 2.5 Å. A higher resolution data set (Native2, Table II) was collected for crystals grown in ammonium phosphate and used in the final refinement. The data were subjected to rigid body refinement, followed by positional and thermal factor ($xyzb$) refinement to 1.93 Å. The final model of the apo UPRTase consists of residues 21–244 of each of the four subunits, eight phosphate ions (two per subunit) and 420 water molecules (Table II). Electron density for residues 1–20 was not found for any monomer, indicating that they are disordered. N-terminal sequencing revealed that the N-terminal methionine was cleaved.

To obtain the UPRTase–uracil, UPRTase–UMP and UPRTase–5-fluorouracil complexes, crystals of the apo UPRTase were soaked with 10 mM uracil for 1 day, 10 mM UMP for 5 days and saturating amounts of 5-fluorouracil for 16 h. Data for the UPRTase–UMP and UPRTase–5-fluorouracil complexes were collected on the ADSC area detector, and the UPRTase–uracil data were collected on the R-AXIS IV. The apo UPRTase, after solvent and phosphate ion removal, was used as the starting model for all complex structures.

Rigid body refinement of the UPRTase–uracil complex dropped the R -factor to 23.4%. Fourier difference maps clearly revealed the location of the bound uracil in all four protein molecules. Subsequently $xyzb$ refinement was carried out. The addition of 258 solvent molecules and eight phosphates (the same phosphates as observed in the apo structure) resulted in an R -factor of 18.1% using data from 10.0 to 2.2 Å resolution (Table II).

The R -factor for the UPRTase–UMP complex dropped to 23.9% after rigid body refinement. After $xyzb$ refinement, electron density was clear for the UMP molecules in each of the four subunits. The four UMP molecules, eight phosphates and 217 solvent molecules were positioned,

Table II. Refinement statistics

	Native2	Uracil	UMP	5-fluorouracil
Resolution range (Å)	10–1.93	10–2.20	10–2.50	10–2.30
No. of reflections	75 978	53 303	32 670	62 830
No. of measurements	178 398	111 427	53 734	91 812
Completeness (%)	92	87	81	86
$I/\sigma(I)$ for all data	11.8	18.5	13.3	6.7
$I/\sigma(I)$ for highest shell				
(2.07–1.93)	2.4	–	–	–
(2.23–2.16)	–	3.9	–	–
(2.85–2.50)	–	–	5.4	–
(2.49–2.26)	–	–	–	2.1
R_{sym}^a	5.7	5.2	3.3	5.2
R -factor ^b	17.9	18.1	14.6	16.4
No. of atoms	7152	7152	7152	7152
No. of solvent molecules	413	288	217	232
No. of phosphate or sulfate ions	8	8	8	12
r.m.s. deviations				
Bond length (Å)	0.014	0.015	0.019	0.018
Bond angles (°)	1.86	1.88	1.99	2.06
Ramachandran plot analysis (%)				
Most favorable	91.1	87.2	88.1	89.1
Allowed	9.3	12.1	11.6	10.3
Generously allowed	0.4	0.8	0.4	0.6
Disallowed	0.0	0.0	0.0	0.0

^a $R_{\text{sym}} = \sum |I_o - \langle I \rangle| / I_o$, where I_o = observed intensity and $\langle I \rangle$ = average intensity obtained from multiple observations of symmetry-related reflections.

^b R -factor = $\sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$; r.m.s.d bond lengths and r.m.s.d bond angles are the root-mean-square deviations from their respective ideal values.

and continued refinement dropped the R -factor to 14.6% using data from 10.0 to 2.5 Å resolution (Table II).

Unlike the crystals used in soaks to prepare the UPRTase–UMP and UPRTase–uracil complexes, those used to obtain the UPRTase–5-fluorouracil complex were grown from solutions of ammonium sulfate. The UPRTase–5-fluorouracil complex was solved as described above for the uracil and UMP complexes. The R -factor was 22.9% after rigid body refinement. After $xyzb$ refinement, the 5-fluorouracil molecules were placed in each of the four subunits, and 232 solvent molecules and 12 sulfates were then added. Continued refinement dropped the R -factor to 16.4% using data from 10.0 to 2.3 Å resolution (Table II).

The current models of the apo, uracil-bound, UMP-bound and 5-fluorouracil-bound structures display excellent stereochemistry as determined by PROCHECK (Laskowski *et al.*, 1993) and display no Ramachandran outliers (Table II). The coordinates of each of the described structures have been deposited in the Brookhaven Protein Data Bank.

Acknowledgements

The authors wish to acknowledge the early scientific contributions of Dr Robert G.K. Donald to this project, and to thank Ms Erika Dahl for preparation of the C128V mutant. This work was supported by a Damon Runyon-Walter Winchell Postdoctoral Fellowship (M.A.S.), the Summerville Fund and Oregon Health Sciences Foundation (R.G.B.), and Public Health Services Grants to R.G.B. (GM-55501) and to B.U. and D.S.R. (AI-31808). D.S.R. and B.U. are Burroughs Wellcome Fund Scholars in Molecular Parasitology and this work is in part supported by a grant from the Burroughs Wellcome Fund.

References

- Argos, P., Hanei, M., Wilson, J.M. and Kelley, W.N. (1983) A possible nucleotide-binding domain in the tertiary fold of phosphoribosyltransferases. *J. Biol. Chem.*, **258**, 6450–6457.
- Asai, T., Lee, C.S., Chandler, A. and O'Sullivan, W.J. (1990) Purification and characterization of uracil phosphoribosyltransferase from *Criethidia luciliae*. *Comp. Biochem. Physiol.*, **95B**, 159–163.
- Baker, E.N. (1995) Solvent interactions with proteins as revealed by X-ray crystallographic studies. In Gregory, R.B. (ed.), *Protein–Solvent Interactions*. Dekker, Inc., New York, pp. 143–189.
- Berens, R.L., Krug, E.C. and Marr, J.J. (1995) Purine and pyrimidine metabolism. In Marr, J.J. and Müller, M. (eds), *Biochemistry and Molecular Biology of Parasites*. Academic Press, London, pp. 150–156.
- Brooks, R.G., Remington, J.S. and Luft, B.J. (1987) Drugs used in the treatment of toxoplasmosis. *Antimicrob. Agents Annu.*, **2**, 297–306.
- Carter, D., Donald, R.G.K., Roos, D. and Ullman, B. (1997) Expression, purification, and characterization of uracil phosphoribosyltransferase from *Toxoplasma gondii*. *Mol. Biochem. Parasitol.*, **87**, 137–144.
- Donald, R.G.K. and Roos, D.S. (1995) Insertional mutagenesis and marker rescue in a protozoan parasite: cloning of the uracil phosphoribosyltransferase locus from *Toxoplasma gondii*. *Proc. Natl Acad. Sci. USA*, **92**, 5749–5753.
- Eads, J.C., Scapin, G., Xu, Y., Grubmeyer, C. and Sacchettini, J.C. (1994) The crystal structure of human hypoxanthine–guanine phosphoribosyltransferase with bound GMP. *Cell*, **78**, 325–334.
- Eads, J.C., Ozturk, D., Wexler, T.B., Grubmeyer, C. and Sacchettini, J.C. (1997) A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase. *Nature Struct. Biol.*, **5**, 47–58.
- Frey, M. (1993) Water structure of crystallized proteins: high resolution studies. In Westhof, E. (ed.), *Water and Biological Macromolecules*. CRC Press, Inc., Boca Raton, FL, pp. 98–147.
- Furey, W. and Swaminathan, S. (1997) PHASES-95: a program package for processing and analyzing diffraction data from macromolecules. *Methods Enzymol.*, **277**, 590–620.
- Goitein, R.K., Chelsky, D. and Parsons, S.M. (1978) Primary ¹⁴C and a secondary ³H substrate kinetic isotope effects for some phosphoribosyltransferases. *J. Biol. Chem.*, **253**, 2963–2971.
- Henriksen, A., Aghajari, N., Jensen, K.F. and Galhede, M. (1996) A flexible loop at the dimer interface is a part of the active site of the adjacent monomer of *Escherichia coli* orotate phosphoribosyltransferase. *Biochemistry*, **35**, 3803–3809.
- Hosaka, H., Nakagawa, A., Tanaka, I., Harada, N., Sano, K., Kimura, M., Yao, M. and Wakatsuki, S. (1997) Ribosomal protein S7: a new RNA-binding motif with structural similarities to a DNA architectural factor. *Structure*, **5**, 1199–1208.
- Iltzsch, M.H. and Tankersley, K.O. (1994) Structure activity relationship of ligands of uracil phosphoribosyltransferase from *Toxoplasma gondii*. *Biochem. Pharmacol.*, **48**, 781–792.

- Jardim,A. and Ullman,B. (1997) The conserved serine-tyrosine dipeptide in *Leishmania donovani* hypoxanthine-guanine phosphoribosyltransferase is essential for catalytic activity. *J. Biol. Chem.*, **272**, 8967-8973.
- Joachimiak,A., Haran,T.E. and Sigler,P.B. (1994) Mutagenesis supports water mediated recognition in the *trp* repressor-operator system. *EMBO J.*, **13**, 367-372.
- Jones,T.A. (1985) Interactive computer program graphics: FRODO. *Methods Enzymol.*, **B115**, 157-171.
- Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110-119.
- Krahn,J., Kim,J.H., Burns,M.R. Parry,R.J., Zalkin,H. and Smith,J. (1997) Coupled formation of an amidotransferase interdomain ammonia channel and a phosphoribosyltransferase active site. *Biochemistry*, **36**, 11061-11068.
- Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of structures. *J. Appl. Crystallogr.*, **24**, 946-950.
- Laskowski,R.A., MacArthur,M.W. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283-291.
- Luft,B.J. and Remington,J.S. (1992) Toxoplasmic encephalitis in AIDS. *Clin. Infect. Dis.*, **15**, 211-222.
- Matthews,B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491-497.
- McIvor,R.S., Wohlhueter,R.M. and Plagemann,P.G.W. (1983) Uracil phosphoribosyltransferase from *Acholeplasma laidlawii*: partial purification and kinetic properties. *J. Bacteriol.*, **156**, 192-197.
- Muchmore,C.R., Krahn,J.M., Kim,J.H., Zalkin,H. and Smith,J.L. (1998) Crystal structure of glutamine phosphoribosylphosphate amidotransferase from *Escherichia coli*. *Protein Sci.*, **7**, 39-51.
- Musick,D.L. (1981) Structural features of the phosphoribosyltransferases and their relationship to the human deficiency disorders of purine and pyrimidine metabolism. *CRC Crit. Rev. Biochem.*, **11**, 1-33.
- Natalini,P., Ruggieri,S., Santarelli,L., Vita,A. and Magni,G. (1979) Baker's yeast UMP:pyrophosphate phosphoribosyltransferase. Purification, enzymatic and kinetic properties. *J. Biol. Chem.*, **254**, 1558-1563.
- Newman,M., Safo,M., Frazoa,C., Khan,G., Zdanov,A., Tickle,I.J., Blundell,T.L. and Andreeva,N. (1991) X-ray analyses of aspartic proteinases: structure and refinement at 2.2 Å resolution of bovine chymotrypsin. *J. Mol. Biol.*, **221**, 1295-1309.
- Nicholls,A., Sharp,K. and Honig,B.H. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281-296.
- Otwinski,Z., Schevitz,R.W., Zhang,R.-G., Lawson,C.L., Joachimiak,A., Marmorstein,R.Q., Luisi,B.F. and Sigler,P.B. (1988) Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature*, **334**, 321-329.
- Ozturk,D.H., Dorfman,R.H., Scapin,G., Sacchettini,J.C. and Grubmeyer,C. (1995) Structure and function of *Salmonella typhimurium* orotate phosphoribosyltransferase: protein complementation reveals shared active sites. *Biochemistry*, **34**, 10764-10770.
- Pfefferkorn,E.R. and Pfefferkorn,L.C. (1977) *Toxoplasma gondii*: characterization of a mutant resistant to 5-fluorodeoxyuridine. *Exp. Parasitol.*, **42**, 44-55.
- Pfefferkorn,E.R., Eckel,M.E.M. and McAdams,E. (1989) The biochemical basis of resistance to emimycin. *Exp. Parasitol.*, **69**, 129-139.
- Rasmussen,U.B., Mygind,B. and Nygaard,P. (1986) Purification and some properties of uracil phosphoribosyltransferase from *Escherichia coli* K12. *Biochim. Biophys. Acta*, **881**, 268-275.
- Rice,P.A., Yang,S., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHF-DNA complex: a protein induced U-turn. *Cell*, **87**, 1295-1306.
- Safo,M.K., Yang,W.Z., Corselli,C., Cramton,S.E., Yuan,H.S. and Johnson,R.C. (1997) The transactivation region of the Fis protein that controls site-specific DNA inverse contains extended mobile β -hairpin arms. *EMBO J.*, **16**, 6860-6873.
- Scapin,G., Ozturk,D.H., Grubmeyer,C. and Sacchettini,J.C. (1995) The crystal structure of the orotate phosphoribosyltransferase complexed with orotate and α -D-5-phosphoribosyl-1-pyrophosphate. *Biochemistry*, **34**, 10744-10754.
- Schumacher,M.A., Carter,D., Roos,D., Ullman,B. and Brennan,R.G. (1996) Crystal structures of *Toxoplasma gondii* HGXPRTase reveal the catalytic role of a long flexible loop. *Nature Struct. Biol.*, **3**, 881-887.
- Schumacher,M.A., Glasfeld,A., Zalkin,H. and Brennan,R.G. (1997) The X-ray structure of the PurR-guanine-*purF* operator complex reveals the contributions of complementary electrostatic surfaces and a water-mediated hydrogen bond to corepressor specificity and binding affinity. *J. Biol. Chem.*, **272**, 22648-22653.
- Schwartzman,J.D. and Pfefferkorn,E.R. (1977) Pyrimidine synthesis by intracellular *Toxoplasma gondii*. *J. Parasitol.*, **67**, 150-156.
- Smith,J.L. (1995) Enzymes of nucleotide synthesis. *Curr. Opin. Struct. Biol.*, **5**, 752-757.
- Smith,J.L. *et al.* (1994) The structure of the allosteric regulatory enzyme of purine biosynthesis. *Science*, **264**, 1427-1433.
- Somoza,J.R., Chin,M.S., Focia,P.J., Wang,C.C. and Fletterick,R.J. (1996) Crystal structure of the hypoxanthine-guanine-xanthine phosphoribosyltransferase from the protozoan parasite *Tritrichomonas foetus*. *Biochemistry*, **35**, 7032-7040.
- Stewart,D.E., Sarkar,A. and Wampler,J.E. (1990) Occurrence and role of *cis* peptide bonds in protein structures. *J. Mol. Biol.*, **214**, 253-260.
- Tamaka,I., Appelt,K., Dijk,J., White,S.W. and Wilson,K.S. (1984) *Nature*, **310**, 376-381.
- Tao,W., Grubmeyer,C. and Blanchard,J.S. (1996) Transition state structure of *Salmonella typhimurium* orotate phosphoribosyltransferase. *Biochemistry*, **35**, 14-21.
- Tronrud,D.E., TenEyck,L.F. and Matthews,B.W. (1985) An efficient general purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr.*, **A43**, 489-501.
- Vis,H., Mariani,M., Vorgias,C.E., Wilson,K.S., Kaptein,R. and Boelens,R. (1995) Solution structure of the HU protein from *Bacillus stearothermophilus*. *J. Mol. Biol.*, **234**, 692-713.
- Vos,S., de Jersey,J. and Martin,J.L. (1997) Crystal structure of *Escherichia coli* xanthine phosphoribosyltransferase. *Biochemistry*, **36**, 4125-4134.
- Weiss,M.S., Metzner,H.J. and Hilgenfeld,R. (1998) Two non-proline *cis* peptide bonds may be important for factor XIII function. *FEBS Lett.*, **423**, 291-296.
- Wimberly,B.T., White,S.W. and Ramakrishnan,V. (1997) The structure of ribosomal protein S7 at 1.9 Å resolution reveals a β -hairpin motif that binds double-stranded nucleic acids. *Structure*, **5**, 1187-1198.
- Xuong,N.H., Nielson,C., Hamlin,R. and Anderson,D.J. (1985) Strategy for data collection from protein crystals using a multiwire counter area detector diffractometer. *J. Appl. Crystallogr.*, **18**, 342-350.

Received March 5, 1998; revised April 14, 1998;
accepted April 15, 1998