

# Molecular Basis of the Antimutagenic Activity of the House-Cleaning Inosine Triphosphate Pyrophosphatase RdgB from *Escherichia coli*

Alexei Savchenko<sup>1</sup>, Michael Proudfoot<sup>1</sup>, Tatiana Skarina<sup>1</sup>,  
Alexander Singer<sup>1</sup>, Olga Litvinova<sup>1</sup>, Ruslan Sanishvili<sup>2</sup>, Greg Brown<sup>1</sup>,  
Nickolay Chirgadze<sup>3</sup> and Alexander F. Yakunin<sup>1\*</sup>

<sup>1</sup>*Banting and Best Department of Medical Research and Ontario Center for Structural Proteomics, University of Toronto, 112 College Street, Toronto, Ontario, Canada M5G 1L6*

<sup>2</sup>*Biosciences Division, Argonne National Laboratory, Argonne, IL 60439, USA*

<sup>3</sup>*Department of Pharmacology, University of Toronto, University Health Network, 200 Elizabeth Street, Toronto, Ontario, Canada M5G 2C4*

Received 20 August 2007;  
received in revised form  
3 October 2007;  
accepted 4 October 2007  
Available online  
11 October 2007

Edited by K. Morikawa

Inosine triphosphate pyrophosphatases, which are ubiquitous house-cleaning enzymes, hydrolyze noncanonical nucleoside triphosphates (inosine triphosphate (ITP) and xanthosine triphosphate (XTP)) and prevent the incorporation of hypoxanthine or xanthine into nascent DNA or RNA. Here we present the 1.5-Å-resolution crystal structure of the inosine triphosphate pyrophosphatase RdgB from *Escherichia coli* in a free state and in complex with a substrate (ITP + Ca<sup>2+</sup>) or a product (inosine monophosphate (IMP)). ITP binding to RdgB induced a large displacement of the  $\alpha$ 1 helix, closing the enzyme active site. This positions the conserved Lys13 close to the bridging oxygen between the  $\alpha$ - and  $\beta$ -phosphates of the substrate, weakening the P <sub>$\alpha$</sub> -O bond. On the other side of the substrate, the conserved Asp69 is proposed to act as a base coordinating the catalytic water molecule. Our data provide insight into the molecular mechanisms of the substrate selectivity and catalysis of RdgB and other ITPases.

© 2007 Elsevier Ltd. All rights reserved.

**Keywords:** inosine triphosphate pyrophosphatase; noncanonical nucleotides; house-cleaning enzyme; crystal structure; site-directed mutagenesis

\*Corresponding author. E-mail address:  
a.iakounine@utoronto.ca.

Abbreviations used: d, deoxy; dITP, deoxyinosine triphosphate; dXTP, deoxyxanthosine triphosphate; dCDP, deoxycytidine diphosphate; dCTP, deoxycytidine triphosphate; dUDP, deoxyuridine diphosphate; dUTP, deoxyuridine triphosphate; GMP, guanosine monophosphate; GTP, guanosine triphosphate; IMP, inosine monophosphate; ITP, inosine triphosphate; MAD, multiple-wavelength anomalous dispersion; NTP, nucleoside triphosphate; PEG, polyethylene glycol; SeMet, selenomethionine; TLC, thin-layer chromatography; XTP, xanthosine triphosphate.

## Introduction

Noncanonical nucleoside triphosphates (NTPs; such as deoxyinosine triphosphate (dITP), deoxyxanthosine triphosphate (dXTP), 8-oxo-deoxyguanosine triphosphate, and 2-oxo-deoxyadenosine triphosphate) are produced from oxidation, deamination, or other modifications of canonical nucleotides.<sup>1–3</sup> DNA or RNA polymerases can incorporate noncanonical (deoxy)NTPs (dNTPs) into nascent DNA or RNA during replication or transcription, resulting in complete blockage of the polymerization reaction or in mispairing with incorrect nucleotides. Removal of noncanonical dNTPs from the cellular nucleotide pool is carried out by “house-cleaning” enzymes whose function is

“to cleanse the cell of potentially deleterious endogenous metabolites and to modulate the accumulation of intermediates in biochemical pathways.”<sup>1,4</sup> Four known families of house-cleaning enzymes include Nudix hydrolases (which hydrolyze various nucleoside diphosphates), trimeric dUTPases (which are specific to dUTP), all  $\alpha$ -NTP pyrophosphatases (which degrade deoxyuridine triphosphate (dUTP), deoxyuridine diphosphate (dUDP), deoxycytidine triphosphate (dCTP), deoxycytidine diphosphate (dCDP), and 2-oxo-deoxyadenosine triphosphate), and ITPases (which degrade dITP and xanthosine triphosphate (XTP)).<sup>1,4,5</sup>

Inosine triphosphate (ITP) is generated by the phosphorylation of inosine monophosphate (IMP), which is a precursor of both AMP and guanosine monophosphate (GMP), whereas dITP can be produced by oxidative deamination of dATP or by reduction of ITP or IDP.<sup>1–3</sup> XTP (dXTP) is formed by oxidative deamination of guanosine triphosphate (GTP) or deoxyguanosine triphosphate (dGTP). If incorporated into DNA, hypoxanthine (from dITP) or xanthine (from dXTP) can be paired with T, C, or A, resulting in potentially deleterious mutations. The incorporation of these noncanonical nucleotides into DNA is prevented by the activity of dITP/XTP pyrophosphatases, which are conserved proteins present in bacteria, archaea, and eukaryotes. Most ITPases belong to the HAM1 family (IPR002637; 682 sequences in databases) named after the hydroxylaminopurine sensitivity protein-1 from yeasts.<sup>6</sup> In *Saccharomyces cerevisiae*, the HAM1 protein protects against the mutagenic effects of the base analog hydroxylaminopurine, which is a natural product of monooxygenase activity on adenine.<sup>6</sup> Only three HAM1 proteins—ITPA from humans, MJ0226 from *Methanococcus jannaschii*, and RdgB from *Escherichia coli*—have been biochemically characterized.<sup>7–10</sup> These proteins showed activity against both canonical and noncanonical nucleotides, but the latter was hydrolyzed 10–100 times more efficiently.<sup>8–10</sup>

Although two ITPase structures have already been published,<sup>7,11</sup> the molecular mechanisms of substrate selectivity and catalysis remain obscure. Structural studies with MJ0226 have demonstrated that the substrate base is exposed to the solvent and has no contacts with the protein.<sup>7</sup> The recent crystal structure of the human ITPA complex with ITP revealed an alternative binding mode for the substrate, with the base sandwiched between two conserved phenylalanines and with the phosphates coordinated by the side chains of several lysines.<sup>11</sup> A similar substrate-binding mode was also suggested by *in silico* substrate-docking experiments with the structure of TM0159, a predicted ITPase.<sup>1</sup>

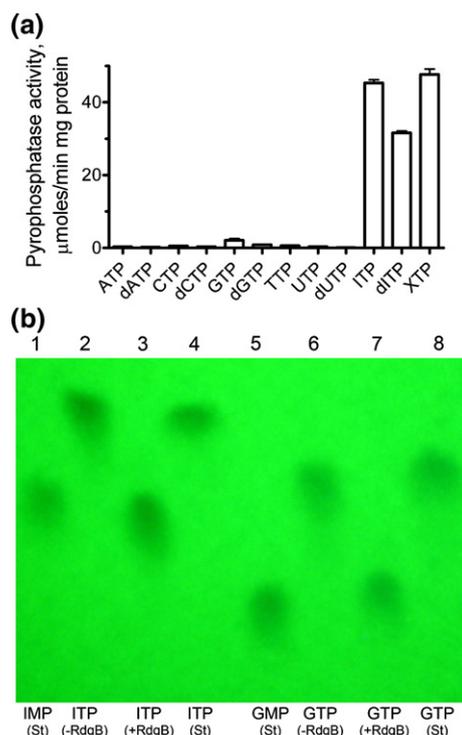
Previous genetic experiments have suggested that *E. coli* RdgB is the enzyme responsible for the interception of dITP/dXTP, preventing the incorporation of hypoxanthine/xanthine into DNA.<sup>12–15</sup> Recent biochemical studies have identified the presence of ITP/XTP pyrophosphatase activity in RdgB, but have produced controversial results on the substrate affinity of this enzyme.<sup>8,10</sup> In this work,

we present the results of structural, biochemical, and mutagenic studies of *E. coli* RdgB and propose a model for its catalytic mechanism. We have identified four amino acid residues (Lys13, Glu41, Lys53, and Asp69) that are absolutely required for catalytic activity, and we have solved the crystal structures of RdgB in a free state or in complex with ITP + Ca<sup>2+</sup> (a substrate) or IMP (a product).

## Results and Discussion

### Purification and oligomeric state

*Escherichia coli* RdgB was overexpressed and affinity purified with a high yield (>30 mg/L culture) of up to >95% homogeneity. Gel-filtration experiments using a Superdex 75 FPLC column showed that the native protein has a molecular mass of 42.3±0.2 kDa, indicating that this protein forms a homodimer in solution. Homologous NTPases from *M. jannaschii* and humans have also been reported to be homodimers in solution.<sup>7,9</sup>



**Fig. 1.** Biochemical characterization of RdgB. (a) Substrate profile: NTP pyrophosphatase activity against different substrates (0.05–1 mM substrate and 7.5 mM Mg<sup>2+</sup>) determined using a pyrophosphatase-coupled assay. (b) Cellulose TLC plate (UV-visualized) showing the reaction products of the hydrolysis of ITP or GTP by RdgB. Samples: (1) IMP standard; (2) the reaction mixture with ITP without RdgB; (3) the reaction mixture with ITP and RdgB; (4) ITP standard; (5) GMP standard; (6) the reaction mixture with GTP without RdgB; (7) the reaction mixture with GTP and RdgB; (8) GTP standard. Experimental conditions were as described in Materials and Methods.

## Enzymatic activity of RdgB

Purified *E. coli* RdgB exhibited a high nucleotide pyrophosphatase activity against three noncanonical NTPs (ITP, dITP, and XTP), a low activity against GTP, and a very low activity against eight other canonical nucleotides (Fig. 1a). RdgB did not discriminate between ribonucleotides and deoxyribonucleotides (Fig. 1a), showed no activity against canonical nucleoside diphosphates or monophosphates, and showed a very low activity with IDP ( $40.1 \pm 3.24$  nmol/min mg protein). Thin-layer chromatography (TLC) analysis demonstrated the formation of IMP during the hydrolysis of ITP and GMP from GTP (Fig. 1b), indicating that RdgB hydrolyzes NTPs to nucleoside monophosphates and pyrophosphate. RdgB had an alkaline pH optimum (pH 9.0) and required the presence of a divalent metal cation for activity.  $Mg^{2+}$  was the most effective metal, whereas  $Co^{2+}$ ,  $Zn^{2+}$ , and  $Mn^{2+}$  supported <50% of the maximum rate (data not shown). In contrast to the human ITPA, which was saturated at 30 mM  $Mg^{2+}$ ,<sup>9</sup> RdgB showed an ~10-times-higher affinity to this metal (apparent  $K_d = 2.52 \pm 0.20$  mM).

**Table 1.** Kinetic parameters of *E. coli* RdgB: wild-type and mutant proteins

Protein	Variable substrate	$K_m$ ( $\mu$ M)	$k_{cat}$ ( $s^{-1}$ )	$k_{cat}/K_m$ ( $M^{-1} s^{-1}$ )
1. Wild type	XTP	23.3 $\pm$ 1.96	19.9 $\pm$ 0.61	$8.54 \times 10^5$
	dITP	11.3 $\pm$ 0.44	13.2 $\pm$ 0.22	$1.17 \times 10^6$
	ITP	5.62 $\pm$ 0.37	18.9 $\pm$ 0.38	$3.38 \times 10^6$
	GTP	312.6 $\pm$ 167.9	0.85 $\pm$ 0.20	$2.72 \times 10^3$
	dGTP	181.5 $\pm$ 23.3	0.37 $\pm$ 0.02	$2.06 \times 10^3$
2. K3A	ITP	289.1 $\pm$ 27.6	0.26 $\pm$ 0.01	$0.90 \times 10^3$
	ITP	5.70 $\pm$ 0.30	16.3 $\pm$ 0.28	$2.86 \times 10^6$
3. T8A	ITP	2.35 $\pm$ 0.15	4.33 $\pm$ 0.07	$1.80 \times 10^6$
4. N10A	ITP	3.04 $\pm$ 0.14	6.42 $\pm$ 0.09	$2.14 \times 10^6$
5. K13A	ITP	nd <sup>a</sup>	nd	nd
6. E16A	ITP	4.05 $\pm$ 0.20	11.3 $\pm$ 0.19	$2.76 \times 10^6$
7. G25T	ITP	16.7 $\pm$ 0.88	24.0 $\pm$ 0.49	$1.41 \times 10^6$
8. E41A	ITP	nd	nd	nd
9. K53A	ITP	11.0 $\pm$ 1.36	0.06 $\pm$ 0.004	$5.46 \times 10^3$
10. D68A <sup>b</sup>	ITP	21.1 $\pm$ 0.88	2.28 $\pm$ 0.03	$1.09 \times 10^5$
11. D69A	ITP	nd	nd	nd
12. S70A	ITP	3.76 $\pm$ 0.26	8.88 $\pm$ 0.18	$2.34 \times 10^6$
13. G71A	ITP	4.28 $\pm$ 0.17	1.26 $\pm$ 0.02	$2.93 \times 10^5$
14. S85A	ITP	7.95 $\pm$ 0.29	26.3 $\pm$ 0.34	$3.29 \times 10^6$
15. R87A	ITP	4.67 $\pm$ 0.47	7.38 $\pm$ 0.22	$1.57 \times 10^6$
16. R114A	ITP	5.94 $\pm$ 0.39	17.1 $\pm$ 0.36	$2.90 \times 10^6$
17. F118A	ITP	9.69 $\pm$ 1.14	16.8 $\pm$ 0.65	$1.73 \times 10^6$
18. C120A	ITP	10.3 $\pm$ 0.90	17.3 $\pm$ 0.53	$1.73 \times 10^6$
19. F154A	ITP	94.9 $\pm$ 31.1	1.09 $\pm$ 0.12	$1.15 \times 10^4$
20. D157A	ITP	4.34 $\pm$ 0.32	11.3 $\pm$ 0.25	$2.63 \times 10^6$
21. F160A	ITP	nd	nd	nd
22. K177A	ITP	23.2 $\pm$ 3.83	8.58 $\pm$ 0.52	$3.70 \times 10^5$
23. S181A	ITP	4.34 $\pm$ 0.32	10.5 $\pm$ 0.16	$2.44 \times 10^6$
24. H182A	ITP	5.93 $\pm$ 0.33	16.4 $\pm$ 0.31	$2.78 \times 10^6$
25. R183A	ITP	119.9 $\pm$ 26.7	4.88 $\pm$ 0.35	$4.07 \times 10^4$

<sup>a</sup> nd, not detected (the specific activity was <10 nmol/min mg protein, except for F160A, which showed significant activity but no saturation).

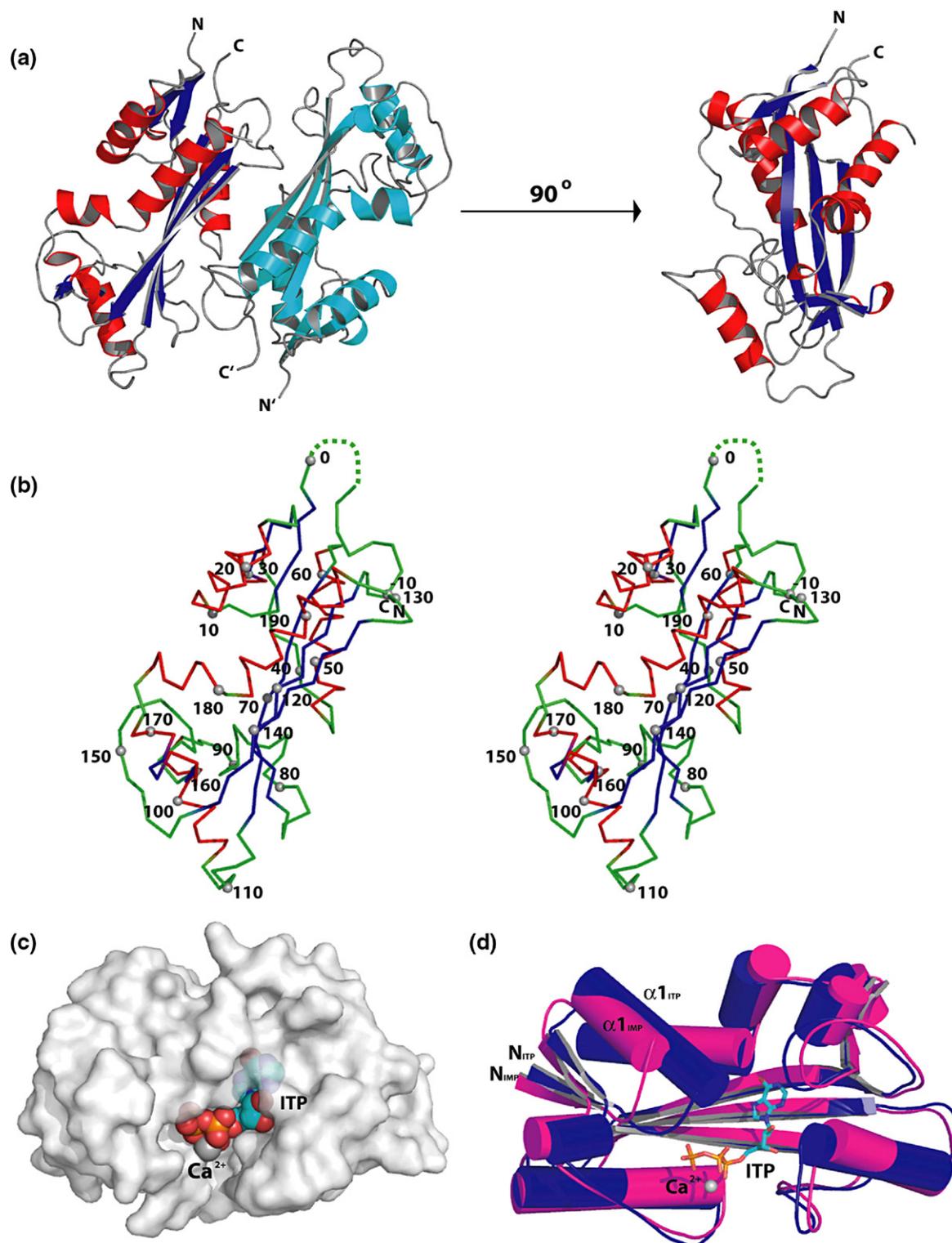
<sup>b</sup> The activity of this protein was measured in the presence of 40 mM  $Mg^{2+}$ , whereas 7.5 mM  $Mg^{2+}$  was used for all other proteins.

With all substrates, RdgB showed sigmoidal saturation kinetics with the Hill coefficient ( $k_H$ ) in the range of 1.7–2.2, suggesting positive cooperativity between two protein subunits in substrate binding (the dimeric state of RdgB was demonstrated by our gel-filtration experiments). In our work, RdgB demonstrated the highest catalytic efficiency ( $k_{cat}/K_m$ ) toward ITP, followed by dITP and XTP (Table 1). Its catalytic efficiency toward noncanonical nucleotides was ~1000 times higher than that toward canonical nucleotides (Table 1). Our results with *E. coli* RdgB (Table 1) and the recent work of Burgis and Cunningham<sup>10</sup> revealed that the affinity of bacterial ITPase to noncanonical NTPs ( $K_m = 5.62$ – $23.3$   $\mu$ M) is 10–100 times higher than the substrate affinities of archaeal (MJ0226) and eukaryotic (ITPA) enzymes ( $K_m = 220$ – $570$   $\mu$ M).<sup>8,9</sup> However, the latter enzymes have higher catalytic activities; therefore, all three ITPases show similar catalytic efficiencies ( $k_{cat}/K_m = 0.61$ – $3.38$   $M^{-1} s^{-1}$ ) toward noncanonical nucleotides. In addition, our *in vitro* experiments revealed that the presence of 0.6–0.8 mM ITP completely suppressed the hydrolysis of 2 mM GTP by RdgB (data not shown). We can assume that, as for the Nudix hydrolase MutT,<sup>16</sup> noncanonical nucleotides might inhibit (overcompete) the hydrolysis of canonical nucleotides by RdgB *in vivo*, making this enzyme even more selective toward noncanonical nucleotides.

## Crystal structure of *E. coli* RdgB in a free state

The crystal structure of *E. coli* RdgB (1k7k) was solved using selenomethionine (SeMet)-substituted protein and multiple-wavelength anomalous dispersion (MAD). The structure demonstrated that the protein is a homodimer (Fig. 2a), consistent with the results of our gel-filtration experiments ( $M_r = 42.3 \pm 0.2$ ). The subunit structure contains one domain with an  $\alpha/\beta$  fold and has 12  $\alpha$ -helices and 6  $\beta$ -strands (Fig. 2b). The crystal structure of RdgB revealed the presence of a large cavity located between the two protein lobes (Fig. 2a and b). Most of the 21 conserved residues of RdgB (including the HAM1 signature motif Ser-His-Arg) are located inside this cleft, suggesting that it accommodates an active site of this enzyme. These residues comprise Thr8, Asn10, Lys13, Glu16, Glu41, Lys53, Asp68, Asp69, Ser70, Gly71, Ser85, Phe154, Asp157, Lys177, Ser181, His182, and Arg183. A similar large cavity holds the enzyme active site in MJ0226 and ITPA.<sup>7,11,17</sup>

A Dali search<sup>17</sup> for RdgB structural homologs identified several strong matches, including the human ITPA (2car; Z-score=7.6; r.m.s.d.=1.87 Å), the putative XTP pyrophosphatase TM0159 from *Thermotoga maritima* (1vp2; Z-score=7.2; r.m.s.d.=1.97 Å), the XTP/ITP pyrophosphatase MJ0226 (1b78; Z-score=6.9; r.m.s.d.=1.77 Å), and the hypothetical protein PH1917 from *Pyrococcus horikoshii* (1v7r and 2dvp; Z-score=5.0; r.m.s.d.=2.36 Å). Structural comparisons also revealed a significant similarity between *E. coli* RdgB YjxX-like



**Fig. 2.** Crystal structure of *E. coli* RdgB. (a) Two views of RdgB (wild type) related by a 90° rotation showing the protein dimer (left) and one subunit (right). The helices, strands, and loop regions of one subunit are in red, blue, and gray, respectively, while the position of the second molecule is shown by helices and strands (both in cyan) and by loops (in gray). The structures show the presence of the large cavity with the  $\beta$ -sheet floor in the center of each subunit. (b) Stereo view of the backbone trace of RdgB. Beginning at residue -10, every 10th residue is highlighted with a gray sphere and labeled; helices are in red, strands are in blue, loops are in green, and disordered regions are shown by a dashed curve. N- and C-termini are also labeled. (c) Surface of the RdgB (D69A) cocrystallized with ITP + Ca<sup>2+</sup>. Note the large pocket with the bound nucleotide and the complete burial of the nucleotide base. The surface has been rendered semitransparent so that the nucleotide base remains visible. (d) Overlay of the apo-RdgB (pink) structure with the structure of the RdgB D69A-ITP-Mg<sup>2+</sup> complex (blue). The positions of N- and C-termini are labeled, and the positions of ITP and metal ion are denoted by a stick diagram and a gray sphere, respectively. The  $\alpha 1$  helix undergoes the largest conformational shift between the two structures.

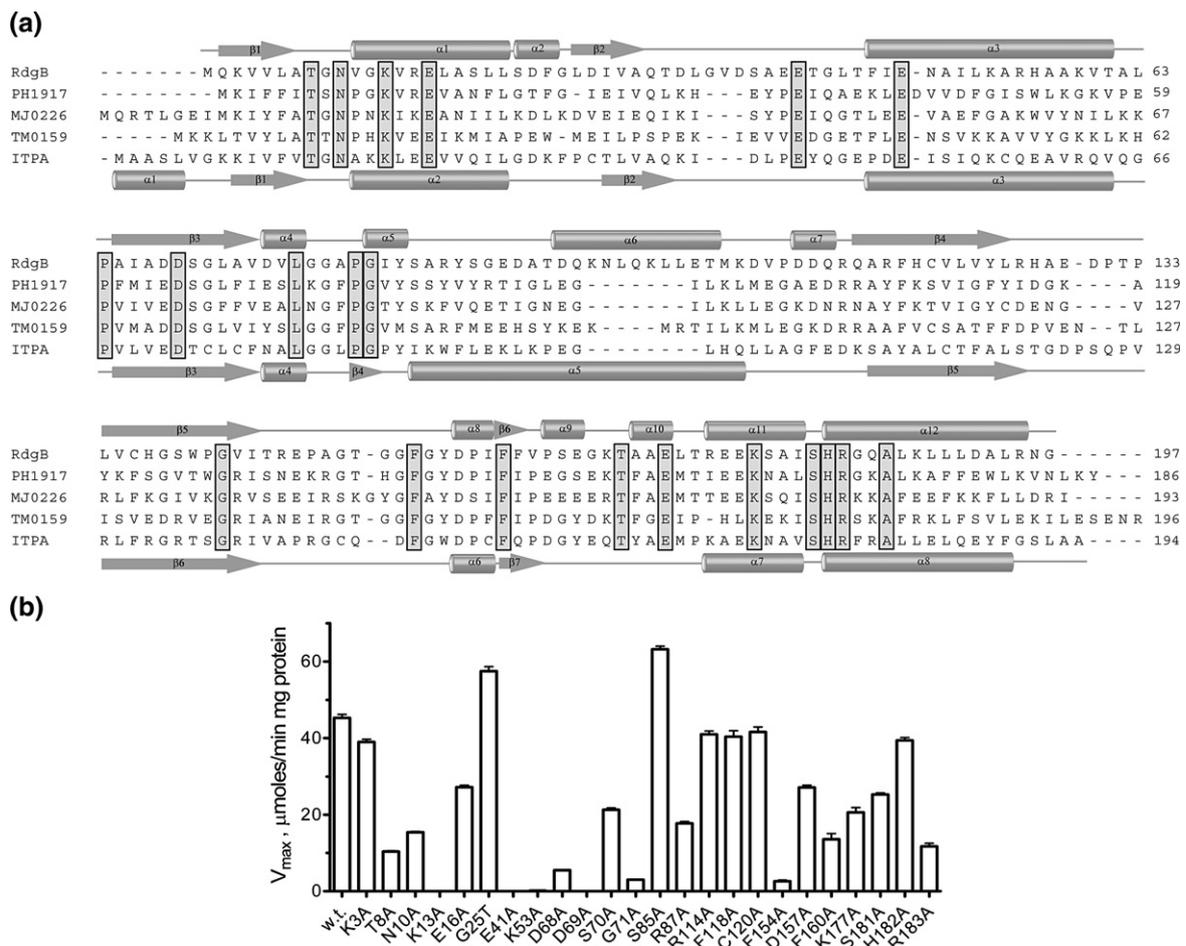
and Maf-like proteins, two other families of predicted nucleotide pyrophosphatases. They include the YjjX proteins from *E. coli* (1u5w; Z-score=4.6)<sup>18</sup> and *Salmonella* (1u14; Z-score=5.1); VC0702 from *Vibrio cholerae* (1zwy; Z-score=5.2); and Maf-like proteins from *Bacillus subtilis* (1ex2; Z-score=4.3)<sup>19</sup> and *Trypanosoma brucei* (2amh; Z-score=4.1). These proteins have a homodimeric organization, similar folds, and catalytic clefts, suggesting a common evolutionary origin among these enzymes. Our experiments with *E. coli* YjjX have revealed the presence of pyrophosphatase activity in ITP, dITP, and XTP (unpublished results), whereas the presence of this activity has not yet been reported in Maf-like proteins.

### Mutational studies of RdgB

Structure-based sequence alignment of RdgB with ITPA and three other HAM1 proteins identified 21 strictly conserved amino acid residues (Fig. 3a). A

BLAST search recognized the presence of the conserved triplet of residues Ser-His-Arg at the C-termini in all sequences of characterized and predicted HAM1-like proteins, which can therefore serve as a signature motif for this family. In the RdgB structure, many conserved residues, including the signature motif Ser-His-Arg, are located within a large cavity that holds the putative active site (Fig. 2a). Site-directed mutagenesis revealed that 13 RdgB mutant proteins showed no effect (Lys3Ala, Gly25Thr, Ser85Ala, Arg114Ala, Phe118Ala, Cys120Ala, and His182Ala) or a small effect (Glu16Ala, Ser70Ala, Arg87Ala, Asp157Ala, Lys177Ala, and Ser181Ala) on activity (Fig. 3b). Seven mutant proteins (Thr8Ala, Asn10Ala, Asp68Ala, Gly71Ala, Phe154Ala, Phe160Ala, and Arg183Ala) demonstrated a greatly reduced ITP pyrophosphatase activity, indicating that these residues are important for catalysis.

Alanine replacement mutagenesis of Lys13, Glu41, Lys53, and Asp69 resulted in a complete loss of enzymatic activity (Fig. 3b). In the RdgB structure,



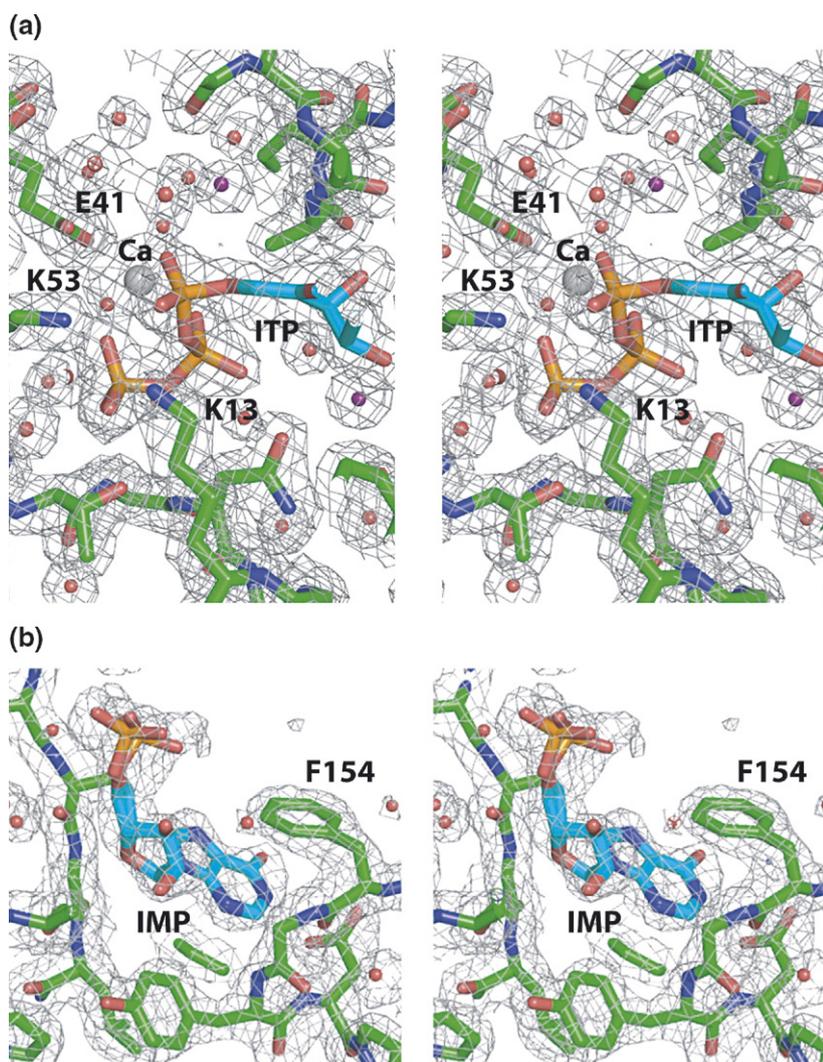
**Fig. 3.** Structure-based sequence alignment of ITPases (a) and site-directed mutagenesis of RdgB (b). (a) Structure-based sequence alignment of RdgB and four other ITPases. Highly conserved residues are highlighted in gray. The secondary structure elements derived from structures of RdgB (1k7k) and ITPA (2car) are shown above and below the alignment, respectively. The compared proteins are *E. coli* RdgB (P52061), PH1917 from *P. horikoshii* (O59580), MJ0226 from *M. jannaschii* (Q57679), TM0159 from *Th. maritima* (Q9WY06), and human ITPA (Q9BY32). (b) Alanine replacement mutagenesis of RdgB: ITP pyrophosphatase activity of purified proteins (0.0013–0.25 mM ITP, 7.5 mM  $\text{Mg}^{2+}$ , and 0.01–1  $\mu\text{g}/\text{assay protein}$ ).

the last three residues create a triad of interacting side chains, Glu41-(3.3 Å)-Lys53-(2.8 Å)-Asp69. Similar triads of interacting lysine and carboxylates can also be recognized in the structures of TM0159 (1vp2) and human ITPA (2car). Asp69 is conserved in all known and predicted ITPases (HAM1 family), as well as in the structurally related Maf proteins and RNase H, where it is predicted to have a catalytic role.<sup>19,20</sup>

All characterized ITPases require a divalent metal cation for activity. It is well known that divalent metal ions usually contribute to enzymatic reactions of nucleotides by coordinating the otherwise flexible phosphate chain.<sup>21,22</sup> We presume that the absence of bound metal in the structure of *E. coli* RdgB and two other free-state structures of ITP pyrophosphatases (MJ0226, TM0159, and ITPA) is due to an inability of the open form of this enzyme to bind the divalent cation. In the structurally similar *E. coli* RNase H,  $Mg^{2+}$  is coordinated by a cluster of four acidic residues (Asp10, Glu48, Asp70, and Asp134).<sup>23</sup> The RdgB Glu41Ala mutant protein was inactive, whereas Glu16Ala and Asp157Ala proteins showed a wild-type affinity to  $Mg^{2+}$  (apparent  $K_d=3.42$  and 1.58 mM, respectively). However, in Asp68Ala

protein, the affinity to  $Mg^{2+}$  was reduced six times (apparent  $K_d=15.6\pm 0.91$  mM), suggesting that this residue is involved (directly or indirectly) in the coordination of  $Mg^{2+}$  in RdgB.

The RdgB Phe118Ala mutant protein showed a 2-times-lower substrate affinity, whereas alanine replacement of the conserved Phe154 reduced the catalytic efficiency ( $k_{cat}/K_m$ ) by almost 300 times (Table 1). In the docking model of TM0159 with bound XTP, the homologous Phe113 and Phe148 were predicted to form a stacking interaction with the substrate base,<sup>1</sup> suggesting the same function for the RdgB Phe118 and Phe154. In contrast to this, in the ITPA structure, the substrate base is sandwiched between closely located Phe149 and Try151,<sup>11</sup> from which only Phe149 is homologous to RdgB Phe154. Interestingly, replacement of the RdgB Gly71 with an alanine residue resulted in a 15-fold reduction in  $k_{cat}$  without an effect on  $K_m$  (Table 1), indicating that the introduction of the small alanine side chain affects the catalytic reaction, perhaps through the effect on the closely located Asp69. This glycine is conserved in all known and predicted ITPases (except for the human ITPA), which contain a



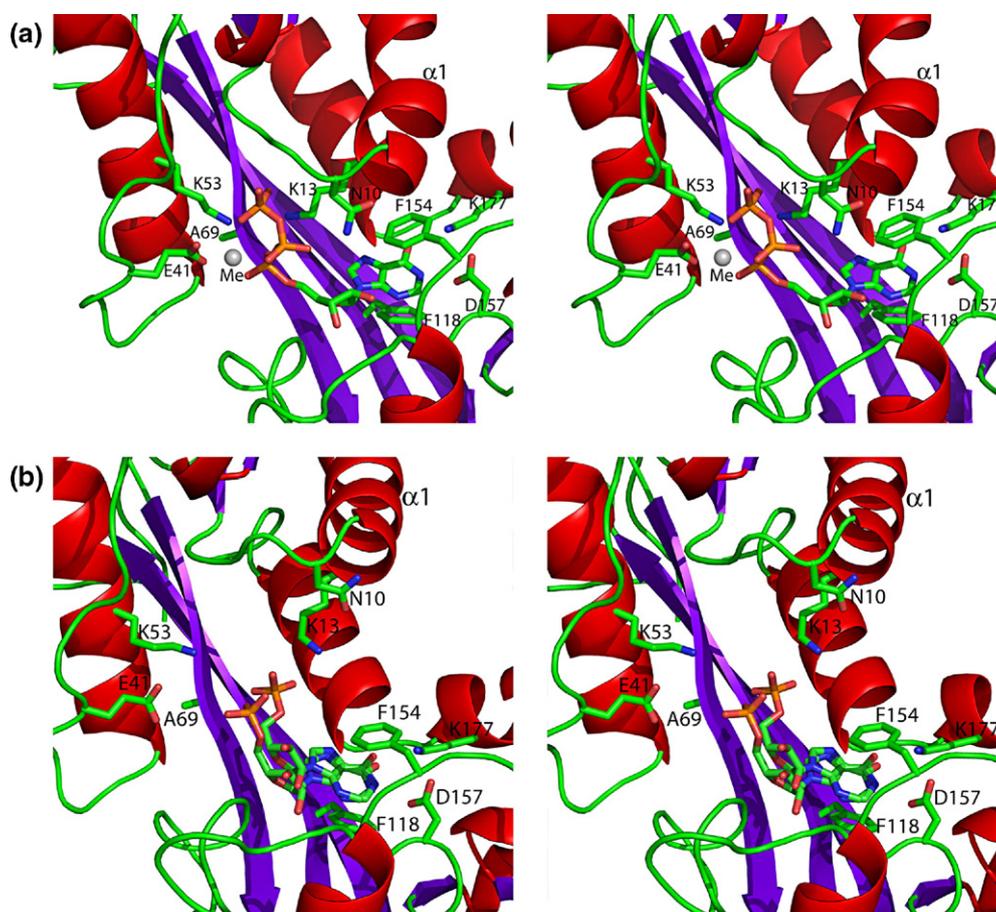
**Fig. 4.** Stereo view of the RdgB catalytic site containing ITP (a) or IMP (b), with  $2F_o - F_c$  electron density contoured at  $1\sigma$ . The positions of ITP, IMP, metal ion (Me), and the side chains of interacting amino acid residues are labeled. In both cases, maps were generated prior to inclusion of the nucleotide in the model. In (a), the map was generated following automated building of the protein model using autoSHARP,<sup>43</sup> while (b) was generated following rigid-body refinement of RdgB protein coordinates.

cysteine residue at this position. In humans, the mutational replacement of nonconserved P32 with threonine results in complete ITPase deficiency in homozygous patients.<sup>24,25</sup> Although ITPase deficiency in humans is not associated with any obvious clinical abnormality, nevertheless, in the long term, it may be an additional genetic factor that predisposes to malignancies or that is responsible for adverse drug reactions.<sup>26–28</sup> However, *E. coli* RdgB contains a glycine residue (Gly25) at this position in sequence alignment (Fig. 3a), and Gly25Thr replacement produces an enzyme with wild-type activity (Fig. 3b).

### Structures of RdgB complexes with ITP + Ca<sup>2+</sup> or IMP

Catalytically inactive RdgB Asp69Ala protein was cocrystallized with ITP in the presence of Ca<sup>2+</sup> (Figs. 2c, 4a, and 5a). The ITP-bound protein crystallized in a different space group (*P*4<sub>2</sub>1<sub>2</sub> as opposed to *P*4<sub>3</sub>2<sub>1</sub>2 for apo-RdgB) (Table 2). In addition, phase information for the ITP-bound protein could not be obtained by molecular replacement with apo-RdgB. Therefore, structure determination of the RdgB+ITP+Ca<sup>2+</sup> complex was done by collecting a single-wavelength anomalous diffraction data set using SeMet-labeled protein.

Analysis of the electron density in this structure confirmed the presence of ITP and Ca<sup>2+</sup> bound to the predicted active site of RdgB (Fig. 4a). The hypoxanthine base of ITP is completely buried within a large pocket (Fig. 2c) and is sandwiched (by  $\pi$ -stacking interactions) between the aromatic side chains of Phe118 (3.2–3.4 Å; bottom of the sandwich) and Phe154 (3.5–3.6 Å; top of the sandwich) (Figs. 5a and 6). Mutational analysis of RdgB (Fig. 3b) revealed that the top aromatic side chain (Phe154) plays a more important role in substrate binding because Phe154Ala mutant showed a 300-times-lower catalytic efficiency, whereas Phe118Ala mutant protein exhibited only a 2-fold increase in *K<sub>m</sub>* (Table 1). In the RdgB-ITP complex, the base is also H-bonded to the side chains of three residues: Asp157 (2.6 Å to the base N1), and Lys177 and Arg183 (2.9 Å to the 6-keto oxygen of hypoxanthine) (Fig. 6). Alanine replacement mutagenesis of Arg183 resulted in a >21-times increase in *K<sub>m</sub>* and a >3-times decrease in activity, whereas Lys177Ala showed a 4-times-lower substrate affinity and a 2-times-lower activity (Table 1), indicating that both these residues play an important role in substrate binding. Both Arg183 and Lys177 contribute to the substrate selectivity of RdgB by discriminating against the 6-amino group of adenine (in ATP) or the 4-amino group of cytosine



**Fig. 5.** Coordination of the substrate and the product in the active site of RdgB. Close-up stereo view of the RdgB D69A active site with bound ITP + Ca<sup>2+</sup> (a) or IMP (b). The metal ion is denoted by a gray sphere and the symbol Me.

**Table 2.** Data collection and refinement statistics

Parameter	apo-RdgB crystal 1			apo-RdgB crystal 2 (1k7k)		RdgB-IMP (2pyu)	RdgB-ITP (2q16)
	Peak	Inflection	Remote	Low resolution	High resolution		
<i>Data collection</i>							
Space group	$P4_32_12$			$P4_32_12$		$P4_32_12$	$P42_12$
Cell dimensions <i>a</i> and <i>c</i> (Å) <sup>a</sup>	<i>a</i> =78.2, <i>c</i> =80.4			<i>a</i> =78.2, <i>c</i> =80.4		<i>a</i> =77.9, <i>c</i> =81.1	<i>a</i> =120.6, <i>c</i> =66.2
Wavelength	0.97900	0.97937	0.95372	1.03321	1.03321	1.54178	0.97942
Resolution (Å)	50–1.79	50–1.79	50–1.82	50–2.61	3.5–1.45	50–2.02	50–1.95
$R_{\text{sym}}$ or $R_{\text{merge}}$	0.065 (0.77 <sup>b</sup> )	0.07 (0.84)	0.07 (0.84)	0.044 (0.015)	0.06 (0.69)	0.0579 (0.476)	0.089 (0.551)
$I/\sigma I$	34.8 (1.9)	32.6 (1.7)	33.5 (0.84)	30.2 (9.7)	20.4 (1.4)	37.5 (5.5)	39.9 (5.5)
Completeness (%)	99.6 (96.9)	99.5 (96.4)	99.6 (96.6)	99.4 (99.7)	98.5 (90.7)	99.9 (99.6)	99.96 (100)
Redundancy	9.5	9.4	9.5	4.0	4.8	4.9 (4.7)	19.4 (17.7)
<i>Refinement</i>							
Resolution (Å)				50–1.5	1.58–1.5	32.03–2.02	26.98–1.95
Number of reflections				37,399	5784	16,119	34,514
$R_{\text{work}}/R_{\text{free}}$				20.0/22.9	22.8/24.4	19.2/25.3	17.1/20.8
Number of atoms							
Protein				1701		1569	2943
Major ligand						46 (IMP, 2 conformations)	64 (Ca-ITP)
Other ligands/ion						4	10
Solvent				255		147	410
<i>B</i> -factors				16.8		32.3	31.5
r.m.s.d.							
Bond lengths (Å)				0.02		0.026	0.015
Bond angles (°)				1.95		2.24	1.68
<i>Ramachandran plot</i>							
Most favored (%)				91.4		90.9	93.1
Additionally allowed (%)				8.0		9.5	6.9
Generously allowed (%)				0		0.6	0
Disallowed (%)				0.6		0	0

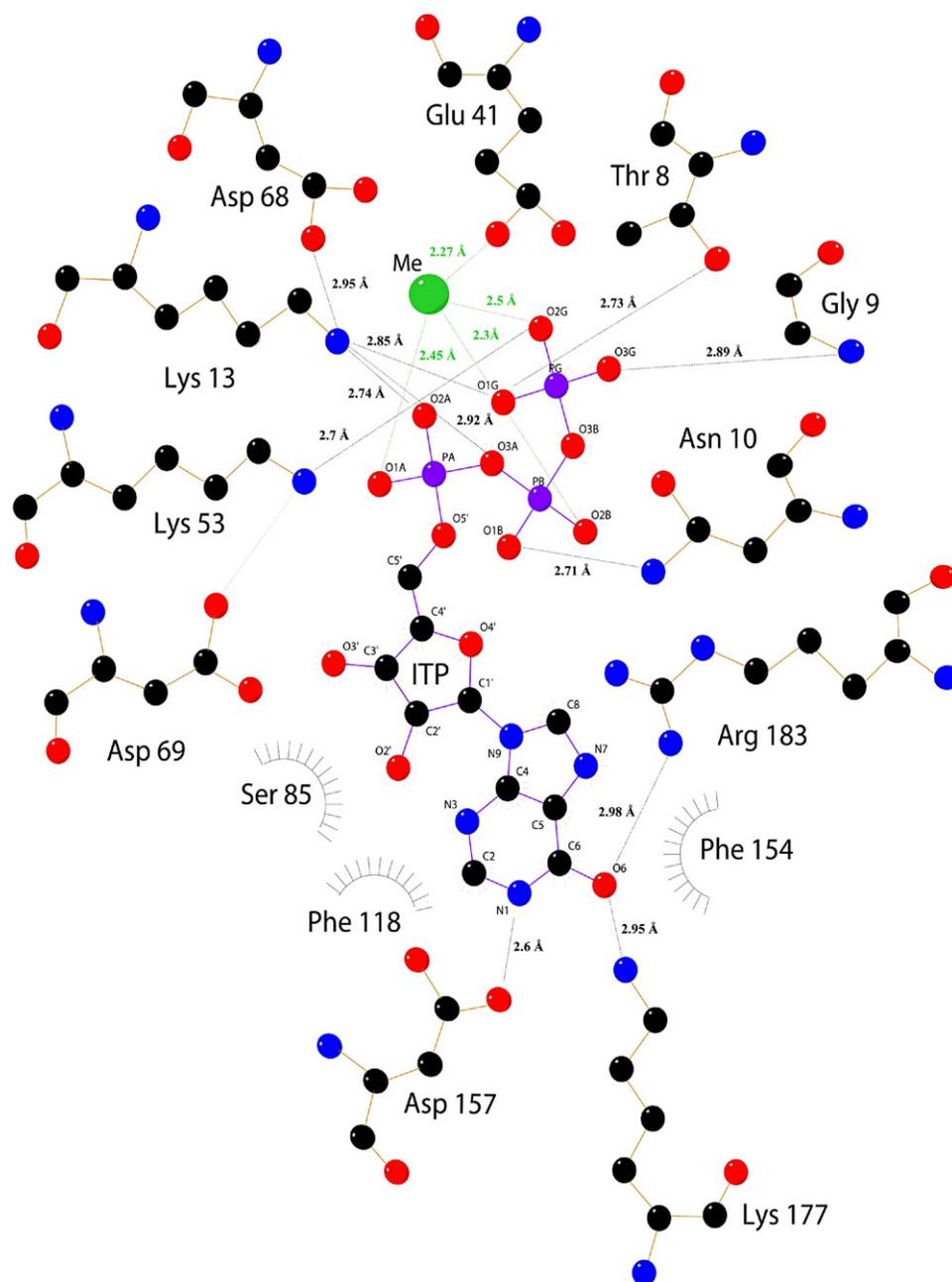
<sup>a</sup> For tetragonal space groups,  $a=b \neq c$  and  $\alpha=\beta=\gamma=90^\circ$ .

<sup>b</sup> Values in parentheses are for the highest-resolution shell.

(in cytidine triphosphate). Additionally, these two residues interact with the side chains of two carboxylates (Lys177-Asp157, 2.6 Å; Arg183-Glu16, 2.9 Å). These interactions are predicted to stabilize the protonated forms of the arginine or lysine side chains<sup>29</sup> that would be important for the substrate discrimination role of Lys177 and Arg183 in RdgB. Asp157 also contributes to substrate coordination through the H-bond to the N1 of the base (2.6 Å) (Fig. 6). The structure suggests that the 2-keto oxygen of xanthine can be accommodated within the buried turn Phe154-Gly155-Tyr156-Asp157, where it can be H-bonded with the main-chain NH groups. The substrate affinity of RdgB to ITP was 4 times higher than the substrate affinity of RdgB to XTP (Table 1), suggesting that the adjustment of the 2-keto oxygen of xanthine in the active site requires more energy. The accommodation of the guanine base in the RdgB active site is hindered by steric interferences of the guanine 2-amino group, with the main-chain amide groups of Phe154-Gly155-Tyr156-Asp157 providing the molecular basis of RdgB discrimination against GTP.

There are no direct contacts between the 2'-hydroxyl or the 3'-hydroxyl group of the ribose ring and the RdgB active site, explaining the absence of discrimination between ribonucleotides and deoxyribonucleotides. Three oxygen atoms of the

triphosphate moiety of ITP are coordinated by interactions with the side chains of three residues: Lys13 (2.8 Å;  $\alpha$ -P), Asn10 (2.7 Å;  $\beta$ -P), Thr8 (2.7 Å;  $\gamma$ -P), and Lys53 (2.7 Å;  $\gamma$ -P), as well as with the main-chain amide groups of Gly9 (2.9 Å;  $\gamma$ -P) and Ser70 (3.2 Å;  $\gamma$ -P) (Fig. 6). In addition, phosphate oxygen atoms chelate the calcium ion with an  $\alpha,\beta,\gamma$  tridentate geometry (each phosphate contributes one oxygen atom to calcium's coordination sphere) (Figs. 5a and 6). The same metal coordination geometry was observed in the crystal structures of the trimeric dUTPase and DNA polymerase  $\beta$ .<sup>21,22</sup> As well, the calcium ion has contacts with the side-chain oxygen of the conserved Glu41 of RdgB (2.05 Å). The recently solved structure of the Maf-like protein Tbru21784aaa from *Tr. brucei* (2amh) revealed a similar coordination of the bound  $Mn^{2+}$  by the side chain of the conserved Glu45 and oxygen atoms of two bound sulfate ions that mimic the  $\alpha$ - and  $\gamma$ -phosphates of the NTP substrate. This similarity suggests some common features in the catalytic mechanism of these enzymes. The described interactions between the substrate and RdgB amino acid residues play an important role in the catalysis, since alanine replacement mutagenesis of these residues produced inactive (Lys13Ala, Glu41Ala, and Lys53Ala) or low-activity (Thr8Ala and Asn10Ala) proteins.



**Fig. 6.** Flat representation of major interactions between the RdgB active site and ITP. Hydrogen bonds are indicated by dashed lines; numbers represent distances in angstroms. The green ball indicates a divalent metal cation ( $\text{Ca}^{2+}$ ). A69 from the RdgB D69A–ITP structure is restored back to aspartate to show the predicted interactions of the D69 side chain with K53 and the substrate.

The recent structure of the human ITPA–ITP complex<sup>11</sup> and the just-released structures of the PH1917 complexes with ITP (2e5x and 2dvo) or IMP (2dvn) revealed similar substrate coordinations in these proteins, with one main difference in the base stacking in ITPA. In the ITPA structure, the base is stacked between the side chains of Phe149 (on top) and Try151 (at the bottom) located one residue apart. In RdgB, the substrate base is sandwiched between the side chains of distantly located phenylalanine (Phe118 at the bottom and Phe154 on top in RdgB). In general, the structures of RdgB, ITPA, and PH1917 agree well with the *in silico* docking model

of the TM1059–XTP complex proposed recently in Galperin *et al.*, and they all provide insights into the mechanism of the substrate selectivity of ITPases.<sup>1</sup>

We have also solved the crystal structure of the catalytically inactive RdgB Asp69Ala mutant cocrystallized with the IMP product (Figs. 4b and 5b). In contrast to RdgB+ITP, the structure of the RdgB+IMP complex was solved by rigid-body refinement using the protein coordinates of apo-RdgB because the unit cell dimensions of the RdgB+IMP crystals and, therefore, the structure of this complex were virtually identical with those of apo-RdgB. The structures of both ITP- and IMP-containing

complexes revealed a similar coordination of the hypoxanthine base in the active site. The phosphate group of IMP was disordered and adopted two conformations with a 3.2-Å distance between the phosphorus atoms (Fig. 5b). The IMP phosphate oxygen atoms interact with the side chains of Lys13 (3.0 Å) and Asp68 (3.4 Å) and with the main-chain NH group of Ser70 (3.1 Å). No electron density could be assigned to  $Mg^{2+}$  in the RdgB-IMP structure. This suggests that, after hydrolysis, the pyrophosphate group with bound  $Mg^{2+}$  exits freely as it makes numerous contacts with the bulk solvent. Thus, in addition to the coordination of the triphosphate moiety of ITP,  $Mg^{2+}$  also facilitates pyrophosphate discharge after catalysis. The overall fold of the RdgB-IMP complex is very similar to the structure of the ligand-free RdgB and represents the protein in an open conformation when the first reaction product (pyrophosphate) has already left the active site, whereas the second product (IMP) is still bound to the enzyme (Fig. 5b). The  $\alpha 1$  helix and the  $\beta 1$ - $\alpha 1$  loop are shifted up by 2.9–7.8 Å to exactly the same position as in the structure of the ligand-free RdgB. The just-released structure of the PH1917 complex with IMP (2dvn) revealed a very similar coordination of the ligand in the active site.

### Substrate-induced conformational changes in RdgB

The superposition of two RdgB structures (free state and ITP complex) revealed that ITP binding induced a large movement of several structural elements toward the bound substrate, resulting in a “closed” conformation of the enzyme (Fig. 2d). The most prominent change includes the downshift of the  $\alpha 1$  helix and the  $\beta 1$ - $\alpha 1$  loop by 2.9 to 7.8 Å, closing the active site and moving a cluster of four conserved residues (Thr8, Asn10, Lys13, and Glu16) and one semiconserved residue (Gly9) close to the substrate (Fig. 2d). In the RdgB-ITP complex structure, the side chain of Lys13 is positioned next to the oxygen of the bridging  $P_{\alpha}$ -O bond of ITP (3.1 Å) (Figs. 5a and 6), and its key role in the catalysis is supported by the complete loss of activity in the Lys13Ala mutant protein (Fig. 3b). The ITP-induced shift of the  $\alpha 1$  helix moves the side chains of the conserved Asn10 and Thr8 close to the oxygen atoms of the substrates  $\beta$ -phosphate (2.7 Å) and  $\gamma$ -phosphate (2.7 Å). Further stabilization of the substrate triphosphate moiety is achieved by the interaction of the main-chain amide group of the semiconserved G9 with the  $\gamma$ -phosphate oxygen. Conserved Glu16 is shifted close to the side chain of the conserved Arg183 (2.9 Å) and reorients the latter residue toward the 6-keto oxygen of the base, moving it by  $\sim 3$  Å in the direction of the substrate. In addition, ITP binding to RdgB downshifted the  $\beta 5$ - $\alpha 8$  loop and the conserved Phe154 by 1.1 Å toward the nucleotide base, locking the substrate in the active site.

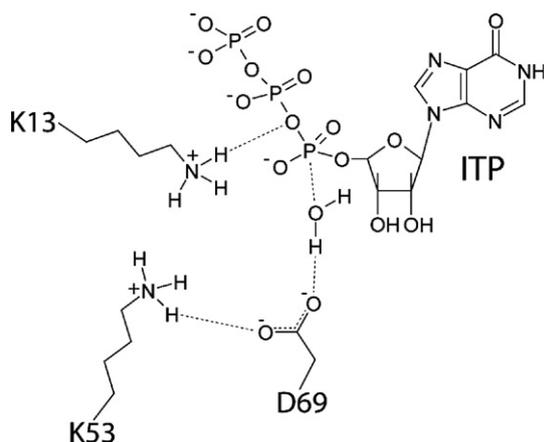
Thus, RdgB structures revealed that this enzyme can have “open” or closed conformations, and their transformations are realized through the concerted

interaction of several static and movable protein elements induced by substrate binding and hydrolysis. The static components include a fixed  $\beta$ -sheet floor accommodating the predicted catalytic residue Asp69 and the associated H-bond network (Lys53 and Glu41), as well as a group of conserved residues involved in substrate binding (Asp68, Ser70, Phe118, His182, and Arg183). The movable elements comprise the  $\alpha 1$  helix and two flexible loops ( $\beta 1$ - $\alpha 1$  and  $\beta 5$ - $\alpha 8$ ) containing residues involved in substrate coordination and catalysis (Thr8, Asn10, Lys13, Glu16, and Phe154). The binding of the ITP- $Ca^{2+}$  complex to the open form of RdgB induces a large shift of movable components toward the bound substrate, transforming the enzyme into a closed conformation, which is reversed back to an open conformation after substrate hydrolysis.

### Implications for the catalytic mechanism of RdgB

Our structural and biochemical data indicate that two networks of interacting lysine and carboxylate residues play an important role in the catalytic mechanism of RdgB: Glu41-Lys53-Asp69 and Lys13-Asp68-Arg183-Glu16. Carboxylate-lysine or carboxylate-arginine interactions stabilize the unprotonated form of carboxylates (by reducing their  $pK_a$ ) or the protonated form of lysine or arginine (by increasing their  $pK_a$ ).<sup>29</sup> These interactions are often found where either a positive charge (from lysine/arginine) or a negative charge (from aspartate/glutamate) is required to bind or polarize a substrate.<sup>29</sup> ITP-induced conformational change brings Asp68 and Lys13 close for interaction (2.9 Å) and establishes the interacting network of four conserved residues (Lys13-Asp68-Arg183-Glu16). This stabilizes the protonated state of Lys13, which is now positioned close (3.1 Å) to the bridging oxygen between the  $\alpha$ - and  $\beta$ -phosphates of ITP, resulting in destabilization of the  $P_{\alpha}$ -O bond (Fig. 7).

On the opposite side of the RdgB active site, the side chain of the conserved Lys53 stabilizes the unprotonated state of Glu41 (which is important for  $Mg^{2+}$  coordination) and Asp69 (a putative catalytic residue; unliganded RdgB structure) (Fig. 7). It has been shown that dUTPases: (1) catalyze the same (pyrophosphatase) reaction on (2) a similar substrate (dUTP); (3) have the same (tridentate) coordination of the metal ion; and (4) use aspartate as a base to activate a water molecule for an in-line nucleophilic attack on the  $\alpha$ -phosphate of dUTP.<sup>21,22,30</sup> These considerations and our structural and mutational studies imply that, in RdgB, the conserved Asp69 is likely to function as a base, which coordinates a catalytic water molecule for the attack on the  $P_{\alpha}$  of ITP (Fig. 7). Asp69 appears to be well positioned to play the role of general base for the buried water molecule. ITP binding to RdgB would shield a bound water molecule from the solvent, promoting the abstraction of a proton to generate a nucleophilic hydroxide ion. When an aspartate side chain is modeled into the position occupied by Ala in an



**Fig. 7.** Proposed catalytic mechanism for RdgB. ITP binding to RdgB leads to trapping of the nucleophilic water molecule in the substrate pocket and its shielding from the surrounding medium. An ITP-induced conformational shift of the RdgB  $\alpha$ 1 helix brings the side chain of K13 close to the  $\alpha$ - $\beta$  bridging oxygen of ITP, enhancing the polarization of this bond and aiding the nucleophilic attack on the  $\alpha$ -P atom. Unprotonated D69 acts as a general base, abstracting a proton from the water molecule and producing a nucleophilic hydroxide ion, which then attacks the  $\alpha$ -phosphate of ITP. The developing negative charge in the  $\alpha$ - $\beta$  bridging oxygen atom appears to be stabilized by a hydrogen bond to the protonated side chain of K13, which then donates its proton to this oxygen atom (on the leaving pyrophosphate group). K13 is then reprotonated by proton transfer from D69 to the side chain of K53, regenerating the active site.

Asp69Ala mutant protein, its side-chain oxygen atoms are positioned 4.3 Å and 5.7 Å from the  $\alpha$ -phosphorus atom of ITP, suggesting that a nucleophilic water molecule might be placed between these atoms. In this structure, there is a water molecule (W1588) located between the ITP  $\alpha$ -phosphate (3.7 Å) and the side-chain oxygen atoms of the modeled Asp69 (2.7 Å and 3.6 Å), which is almost ideally positioned for in-line nucleophilic attack on the  $\alpha$ -phosphate. In the free-state structure (1k7k), two water molecules (W20 and W167) are located close (2.8 Å and 3.0 Å, respectively) to the side-chain oxygen atoms of Asp69, and one of them might play the role of a nucleophilic water molecule. The proton for the leaving O-atom of pyrophosphate might be donated directly by the side chain of protonated Lys13, which can be reprotonated by proton transfer from Asp69 to the side chain of Lys53 (3.9 Å from Lys13) (Fig. 7).

Thus, the *E. coli* ITPase RdgB demonstrates exquisite substrate specificity with respect to the base of noncanonical nucleotides (ITP, dITP, and XTP) that is essential in keeping their intracellular concentration low and in preventing the unintended hydrolysis of canonical nucleotides. Our work provided structural insights into substrate selectivity and proposed a model for the catalytic mechanism of ITPases, which will be examined in future structural studies.

## Materials and Methods

### Gene cloning and protein purification

*Escherichia coli* RdgB (*yggV*) was cloned into the pET15b-based (Stratagene) vector p11 containing the N-terminal 6His affinity tag, TEV cleavage site, and tandem terminator codons.<sup>31</sup> RdgB was overexpressed in *E. coli* BL21 (DE3) and affinity purified using  $\text{Ni}^{2+}$ -chelate chromatography, as previously described.<sup>31</sup>

### Enzyme assays

NTP pyrophosphatase activity was determined by measuring the  $\text{P}_i$  release in a 160- $\mu\text{L}$  assay solution in 96-well microplates containing 50 mM Tris-HCl (pH 9.0), 5 mM  $\text{Mg}^{2+}$ , 1.25–250  $\mu\text{M}$  substrate, and 0.01–0.1  $\mu\text{g}$  RdgB, using 10 mU of inorganic pyrophosphatase (from baker's yeast; Sigma). The nucleotide substrates were purchased from Sigma, TriLink BioTechnologies, and Jena Biosciences. After 20 min of incubation at 37 °C, the reaction was stopped by the addition of 40  $\mu\text{L}$  of malachite green reagent,<sup>32</sup> and absorbance was measured at 630 nm. Kinetic parameters were determined by nonlinear curve fitting using GraphPad Prism software (version 4.00 for Windows; GraphPad Software, San Diego, CA)†. TLC analysis of the RdgB reaction products was performed using cellulose plates (Sigma) and solvent A (saturated ammonium sulfate/3 M sodium acetate/isopropyl alcohol, 80:6:2).<sup>33</sup> The reaction products and nucleotide standards were visualized under UV light.

### Site-directed mutagenesis

The selected residues of the wild-type RdgB were mutated to alanine using PCR with mutagenic primers and the QuikChange mutagenesis kit (Stratagene). The presence of mutations was confirmed by DNA sequencing using the T7 promoter primer. Plasmid mini preps were made using a Qiaprep Spin Mini Prep kit (Qiagen). The 6His-tagged proteins were overexpressed in *E. coli* BL21 (DE3) and affinity purified on a  $\text{Ni}^{2+}$ -chelate column, as previously described.<sup>31</sup> The yield and purity of the proteins were estimated on 12.5% SDS polyacrylamide gels after visualization by Coomassie blue staining.

### Protein crystallization

Protein crystals were grown at 21 °C by the hanging-drop vapor-diffusion method, with 2  $\mu\text{L}$  of protein sample mixed with an equal volume of reservoir buffer. The crystals of wild-type RdgB grew after 3–5 days in the presence of 30% polyethylene glycol (PEG) 4000, 0.2 M ammonium acetate, and 0.1 M Na citrate (pH 5.8). The crystals of the RdgB complexes with IMP or ITP were obtained by crystallization of the RdgB D69A protein in the presence of 20% PEG 3350, 0.2 M Mg acetate, and 10 mM IMP (for the IMP complex), or in the presence of 22.5% PEG 3350, 0.18 M ammonium sulfate, 0.1 M sodium acetate (pH 4.6), 20 mM calcium chloride, 2% isopropanol, and 10 mM ITP (for the ITP complex). For diffraction studies, the crystals were stabilized with the crystallization buffer

† [www.graphpad.com](http://www.graphpad.com)

supplemented with 20% ethylene glycol as a cryoprotectant and flash frozen in liquid nitrogen.

### Structure determination

For the wild-type RdgB crystal, a three-wavelength MAD experiment was carried out with a SeMet-derivatized protein crystal on the 19ID beam line of the Structural Biology Center at the Advanced Photon Source (Argonne, IL). The statistics of data collection and processing are given in Table 2. Diffraction data were collected at the inflection point, at the peak of the selenium absorption edge, and at a remote energy above the edge (Table 2, crystal 1). High-resolution (up to 1.45 Å) diffraction data were collected from crystal 2 (Table 2, crystal 2). Because the dynamic range of crystal diffraction was broader than that of the detector, two scans were required for low and high resolutions. Data collection and visualization were done with d\*TREK,<sup>34</sup> and all data were integrated into and scaled with the program package HKL2000.<sup>35</sup> MAD phasing of the RdgB data was carried out with the program CNS.<sup>36</sup> More than 80% of the main chain and about half of the side chains were built automatically with ARP/wARP.<sup>37</sup> The remainder of the model was built and all side chains were corrected manually using the program O.<sup>38</sup> All subsequent refinements were carried out with REFMAC5<sup>39</sup> within the CCP4 suite of programs.<sup>40</sup> For the RdgB-IMP crystal, the diffraction data were collected on a Rigaku RU-H3R rotating anode equipped with Osmic confocal "blue" optics, and diffraction intensities were recorded on an R-Axis IV<sup>++</sup> image plate (Rigaku Americas, Texas, USA) (Table 2). The structure was solved by using the atomic coordinates of the apo-RdgB structure (1k7k) and subjecting them to rigid-body refinement using REFMAC5<sup>41</sup> within the CCP4 suite version 6.0.2.<sup>42</sup> The diffraction data from the SeMet-derivatized RdgB-ITP crystal were collected on the 19ID beam line at the peak of the selenium absorption edge. Initial phases were obtained by single-wavelength anomalous diffraction phasing using autoSHARP;<sup>43</sup> subsequent phases were subjected to density modification and automatic model building, resulting in >90% of the protein model being built. Additional model building and refinement of the RdgB-ITP and RdgB-IMP structures were accomplished using Coot.<sup>44</sup> Images were generated using PyMol<sup>‡</sup> and LIGPLOT.<sup>45</sup>

### Protein Data Bank accession codes

Coordinates and structure factors have been deposited with accession codes 1k7k (wild-type RdgB; free state), 2pyu (D69A RdgB-IMP complex), and 2q16 (D69Av-ITP complex).

### Acknowledgements

We thank all members of the Ontario Center for Structural Proteomics for help in the conduct of experiments. The authors acknowledge the support of Genome Canada (through the Ontario Genomics Institute), the Ontario Research and Development

Challenge Fund, and the Protein Structure Initiative of the National Institutes of Health (Midwest Center for Structural Genomics, NIH grant GM 62414). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract no. W-31-109-ENG-38.

### References

- Galperin, M. Y., Moroz, O. V., Wilson, K. S. & Murzin, A. G. (2006). House cleaning, a part of good house-keeping. *Mol. Microbiol.* **59**, 5–19.
- Kamiya, H. (2003). Mutagenic potentials of damaged nucleic acids produced by reactive oxygen/nitrogen species: approaches using synthetic oligonucleotides and nucleotides: survey and summary. *Nucleic Acids Res.* **31**, 517–531.
- Nguyen, T., Brunson, D., Crespi, C. L., Penman, B. W., Wishnok, J. S. & Tannenbaum, S. R. (1992). DNA damage and mutation in human cells exposed to nitric oxide *in vitro*. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 3030–3034.
- Bessman, M. J., Frick, D. N. & O'Handley, S. F. (1996). The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J. Biol. Chem.* **271**, 25059–25062.
- Koonin, E. V. (1993). A highly conserved sequence motif defining the family of MutT-related proteins from eubacteria, eukaryotes and viruses. *Nucleic Acids Res.* **21**, 4847.
- Noskov, V. N., Staak, K., Shcherbakova, P. V., Kozmin, S. G., Negishi, K., Ono, B. C. *et al.* (1996). *HAM1*, the gene controlling 6-N-hydroxylaminopurine sensitivity and mutagenesis in the yeast *Saccharomyces cerevisiae*. *Yeast*, **12**, 17–29.
- Hwang, K. Y., Chung, J. H., Kim, S. H., Han, Y. S. & Cho, Y. (1999). Structure-based identification of a novel NTPase from *Methanococcus jannaschii*. *Nat. Struct. Biol.* **6**, 691–696.
- Chung, J. H., Park, H. Y., Lee, J. H. & Jang, Y. (2002). Identification of the dITP- and XTP-hydrolyzing protein from *Escherichia coli*. *J. Biochem. Mol. Biol.* **35**, 403–408.
- Lin, S., McLennan, A. G., Ying, K., Wang, Z., Gu, S., Jin, H. *et al.* (2001). Cloning, expression, and characterization of a human inosine triphosphate pyrophosphatase encoded by the *itpa* gene. *J. Biol. Chem.* **276**, 18695–18701.
- Burgis, N. E. & Cunningham, R. P. (2007). Substrate specificity of RdgB protein, a deoxyribonucleoside triphosphate pyrophosphohydrolase. *J. Biol. Chem.* **282**, 3531–3538.
- Stenmark, P., Kursula, P., Flodin, S., Graslund, S., Landry, R., Nordlund, P. & Schuler, H. (2007). Crystal structure of human inosine triphosphatase. Substrate binding and implication of the inosine triphosphatase deficiency mutation P32T. *J. Biol. Chem.* **282**, 3182–3187.
- Clyman, J. & Cunningham, R. P. (1987). *Escherichia coli* K-12 mutants in which viability is dependent on recA function. *J. Bacteriol.* **169**, 4203–4210.
- Bradshaw, J. S. & Kuzminov, A. (2003). RdgB acts to avoid chromosome fragmentation in *Escherichia coli*. *Mol. Microbiol.* **48**, 1711–1725.
- Burgis, N. E., Brucker, J. J. & Cunningham, R. P. (2003). Repair system for noncanonical purines in *Escherichia coli*. *J. Bacteriol.* **185**, 3101–3110.

‡ <http://www.pymol.org>

15. Lukas, L. & Kuzminov, A. (2006). Chromosomal fragmentation is the major consequence of the *rdgB* defect in *Escherichia coli*. *Genetics*, **172**, 1359–1362.
16. Saraswat, V., Massiah, M. A., Lopez, G., Amzel, L. M. & Mildvan, A. S. (2002). Interactions of the products, 8-oxo-dGMP, dGMP, and pyrophosphate with the MutT nucleoside triphosphate pyrophosphohydrolase. *Biochemistry*, **41**, 15566–15577.
17. Holm, L. & Sander, C. (1994). Searching protein structure databases has come of age. *Proteins: Struct. Funct. Genet.* **19**, 165–173.
18. Zheng, J., Singh, V. K. & Jia, Z. (2005). Identification of an ITPase/XTPase in *Escherichia coli* by structural and biochemical analysis. *Structure*, **13**, 1511–1520.
19. Minasov, G., Teplova, M., Stewart, G. C., Koonin, E. V., Anderson, W. F. & Egli, M. (2000). Functional implications from crystal structures of the conserved *Bacillus subtilis* protein Maf with and without dUTP. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6328–6333.
20. Goedken, E. R. & Marqusee, S. (2001). Co-crystal of *Escherichia coli* RNase HI with  $Mn^{2+}$  ions reveals two divalent metals bound in the active site. *J. Biol. Chem.* **276**, 7266–7271.
21. Chan, S., Segelke, B., Lekin, T., Krupka, H., Cho, U. S., Kim, M. Y. *et al.* (2004). Crystal structure of the *Mycobacterium tuberculosis* dUTPase: insights into the catalytic mechanism. *J. Mol. Biol.* **341**, 503–517.
22. Barabas, O., Pongracz, V., Kovari, J., Wilmanns, M. & Vertessy, B. G. (2004). Structural insights into the catalytic mechanism of phosphate ester hydrolysis by dUTPase. *J. Biol. Chem.* **279**, 42907–42915.
23. Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M. *et al.* (1990). Three-dimensional structure of ribonuclease H from *E. coli*. *Nature*, **347**, 306–309.
24. Sumi, S., Marinaki, A. M., Arenas, M., Fairbanks, L., Shobowale-Bakre, M., Rees, D. C. *et al.* (2002). Genetic basis of inosine triphosphate pyrophosphohydrolase deficiency. *Hum. Genet.* **111**, 360–367.
25. Cao, H. & Hegele, R. A. (2002). DNA polymorphisms in ITPA including basis of inosine triphosphatase deficiency. *J. Hum. Genet.* **47**, 620–622.
26. Vanderheiden, B. S. (1969). Genetic studies of human erythrocyte inosine triphosphatase. *Biochem. Genet.* **3**, 289–297.
27. Holmes, S. L., Turner, B. M. & Hirschhorn, K. (1979). Human inosine triphosphatase: catalytic properties and population studies. *Clin. Chim. Acta*, **97**, 143–153.
28. Derijks, L. J., Gilissen, L. P., Hooymans, P. M. & Hommes, D. W. (2006). Review article: thiopurines in inflammatory bowel disease. *Aliment. Pharmacol. Ther.* **24**, 715–729.
29. Gutteridge, A. & Thornton, J. M. (2005). Understanding nature's catalytic toolkit. *Trends Biochem. Sci.* **30**, 622–629.
30. Larsson, G., Svensson, L. A. & Nyman, P. O. (1996). Crystal structure of the *Escherichia coli* dUTPase in complex with a substrate analogue (dUDP). *Nat. Struct. Biol.* **3**, 532–538.
31. Zhang, R. G., Skarina, T., Katz, J. E., Beasley, S., Khachatryan, A., Vyas, S. *et al.* (2001). Structure of *Thermotoga maritima* stationary phase survival protein SurE: a novel acid phosphatase. *Structure*, **9**, 1095–1106.
32. Baykov, A. A., Evtushenko, O. A. & Avaeva, S. M. (1988). A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal. Biochem.* **171**, 266–270.
33. Genschik, P., Billy, E., Swianiewicz, M. & Filipowicz, W. (1997). The human RNA 3'-terminal phosphate cyclase is a member of a new family of proteins conserved in Eucarya, Bacteria and Archaea. *EMBO J.* **16**, 2955–2967.
34. Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallogr. Sect. D*, **55**, 1718–1725.
35. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr. Part A*, **276**, 307–326.
36. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D*, **54**, 905–921.
37. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463.
38. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. Sect. A*, **47**, 110–119.
39. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D*, **53**, 240–255.
40. Collaborative Computing Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D*, **50**, 760–763.
41. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. Sect. D*, **55**, 247–255.
42. Winn, M. D., Ashton, A. W., Briggs, P. J., Ballard, C. C. & Patel, P. (2002). Ongoing developments in CCP4 for high-throughput structure determination. *Acta Crystallogr. Sect. D*, **58**, 1929–1936.
43. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2006). Automated structure solution with auto-SHARP. *Methods Mol. Biol.* **364**, 215–230.
44. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.
45. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134.