Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structure-based identification of inositol polyphosphate 1-phosphatase from *Entamoeba histolytica*

Khaja Faisal Tarique,[a] Syed Arif Abdul Rehman,[a] Christian Betzel[b] and Samudrala Gourinath[a]*

[a]School of Life Sciences, Jawaharlal Nehru University, New Delhi 110 067, India, and [b]Laboratory for Structural Biology of Infection and Inflammation, Department of Chemistry, c/o DESY, University of Hamburg, Notkestrasse 85, 22603 Hamburg, Germany

Correspondence e-mail: sgourinath@mail.jnu.ac.in, samudralag@yahoo.com

Received 10 May 2014
Accepted 24 September 2014

PDB reference: inositol polyphosphate 1-phosphatase, 4qxd

Inositol polyphosphate 1-phosphatase from *Entamoeba histolytica* (EhIPPase) is an Mg^{2+}-dependent and Li^{+}-sensitive enzyme that catalyzes the hydrolysis of inositol 1,4-bisphosphate [Ins(1,4)P2] into *myo*-inositol 1-monophosphate and PO_4^{3-}. In the present work, EhIPPase has been biochemically identified and its crystal structure has been determined in the presence of Mg^{2+} and PO_4^{3-} at 2.5 Å resolution. This enzyme was previously classified as a 3′(2′),5′-bisphosphate nucleotidase in the NCBI, but its biochemical activity and structural analysis suggest that this enzyme behaves more like an inositol polyphosphate 1-phosphatase. The ability of EhIPPase to hydrolyze the smaller Ins(1,4)P2 better than the bulkier 3′-phosphoadenosine 5′-phosphate (PAP) is explained on the basis of the orientations of amino-acid residues in the binding site. This structure is the first of its class to be determined from any protozoan parasite, and is the third to determined among all organisms, following its rat and bovine homologues. The three-dimensional fold of EhIPPase is similar to those of other members of the inositol monophosphatase superfamily, which also includes inositol monophosphatase, 3′(2′),5′-bisphosphate nucleotidase and fructose-1,6-bisphosphate 1-phosphatase. They all share conserved residues essential for metal binding and substrate hydrolysis, with the motif D-X_n-EE-X_n-DP(I/L)DG(S/T)-X_n-WD-X_n-GG. The structure is divided into two domains, namely $\alpha+\beta$ and α/β, and the substrate and metal ions bind between them. However, the ability of each enzyme class to act specifically on its cognate substrate is governed by the class-specific amino-acid residues at the active site.

1. Introduction

Entamoeba histolytica is an enteric protozoan parasite that causes amoebic dysentery and liver abscesses. Infection usually takes place by consuming food or water contaminated with environmentally stable cysts. Enteric infection begins with excystation and is followed by the release of mononucleate trophozoites (Choudhuri & Rangan, 2012). The disease occurs when trophozoites invade the intestinal lumen, enter into the circulatory system and finally make their way to other vital organs such as the liver, brain and lungs, where they cause abscesses (Salles *et al.*, 2007). Cell motility, phagocytosis, adherence to the host surface and proteolytic degradation are the major functions of virulence factors that enable *E. histolytica* to infect the host cell (Gilchrist & Petri, 1999; Padilla-Vaca & Anaya-Velázquez, 2010; Christy & Petri, 2011). This mode of infection is particularly well mediated by lipid rafts.

Lipid rafts are highly ordered, nonhomogenously distributed microdomains that are found in the plasma membrane of many eukaryotic cells and play an important role in cell–cell

and cell–matrix adhesion, endocytosis and secretion (Brown & London, 1998; Goldston, Powell & Temesvari, 2012). These lipid rafts have also been identified in *E. histolytica* and are found to harbour the Gal/GalNAc lectin, an important virulence factor that plays a role in the interaction of this pathogen with the extracellular matrix of the host (Pacheco *et al.*, 2004; Laughlin *et al.*, 2004). This interaction initiates a plethora of signalling events and the recruitment of many proteins related to pathogenicity (Goldston, Powell, Koushik *et al.*, 2012). These processes are in turn regulated by the phosphatidyl signalling pathway. Phosphatidylinositol (PI) functions as a precursor for the synthesis of signalling phosphoinositides and GPI-anchored proteins in a number of protozoans, including *E. histolytica*, and has been shown to participate in the virulence functions described above (Koushik *et al.*, 2014). One of the important signalling lipids involved in the phosphatidyl signalling pathway is phosphatidylinositol 4,5-bisphosphate [PI(4,5)P2], which in a number of eukaryotic cells acts as a second messenger and binds to and modulates effector proteins such as actin and its binding proteins adhesin, filamin, filopodin, coactosin *etc.* (Gamper & Shapiro, 2007). Interestingly, PI(4,5)P2 has been identified in lipid rafts of polarized trophozoite cells and found to regulate processes such as cytoskeleton remodelling, apoptosis, adhesion and phagocytosis in *E. histolytica* (Koushik *et al.*, 2013). Exposure of trophozoites to cholesterol has been shown to increase their virulence and to enrich PI(4,5)P2 in the lipid rafts (Koushik *et al.*, 2013). Other lipid second messengers generated from PI(4,5)P2, such as inositol 3-phosphate (IP3) and diacylglycerol, activate protein kinase C and increase the intracellular Ca^{2+} levels (Berridge, 1984). Ca^{2+} is a potent stimulator of motility through the activation of myosin light-chain kinase (MLCK), leading to an increase in myosin II-based contractility in *E. histolytica* (Koushik *et al.*, 2013). Moreover, a number of calmodulin-like calcium-binding proteins have been identified in this pathogen and have been found to be involved in cytoskeleton dynamics and scission machinery during erythrophagocytosis (Aslam *et al.*, 2012; Grewal *et al.*, 2013; Kumar *et al.*, 2007; Bhattacharya *et al.*, 2006). Phosphorylated signalling phosphatidylinositol molecules such as phosphatidylinositol 3-phosphate (PI3P), phosphatidylinositol 3,4-bisphosphate [PI(3,4)P2], phosphatidylinositol 3,5-bisphosphate [PI(3,5)P2] and phosphatidylinositol 3,4,5-triphosphate [PI(3,4,5)P3] interact with and regulate a number of effector proteins having FYVE finger domains and pleckstrin homology domains (Powell *et al.*, 2006; Lemmon, 2008). The amount of these phosphoinositides in the plasma membrane of *E. histolytica* is greater than that found in mammalian cells (Koushik *et al.*, 2014).

Biochemical disruption of these lipid rafts through the administration of wortmannin has been shown to result in a dose-dependent decrease in trophozoite adhesion to the host cell and to allay the virulence of the pathogen (López-Contreras *et al.*, 2013; Rivière *et al.*, 2007). The amount and cellular concentration of these phosphoinositides are regulated by a fine interplay of the respective phosphatidylinositol kinases and inositol phosphatases.

In this study, we have structurally and biochemically identified inositol polyphosphate 1-phosphatase from *E. histolytica* (EhIPPase), which is otherwise annotated as a 3′(2′),5′-bisphosphate nucleotidase in the KEGG pathway. The enzyme catalyzes the hydrolysis of the 1-position phosphate from inositol 1,4-bisphosphate in the presence of Mg^{2+} to release *myo*-inositol-4-phosphate, which is then again hydrolysed by inositol monophosphatase to yield *myo*-inositol. *myo*-Inositol is then recycled into the phosphatidyl signalling pathway, serving as a building block for various other inositol phosphates. The enzyme nevertheless showed very poor activity with 3′-phosphoadenosine 5′-phosphate. It is now clear that of the two enzymes annotated as 3′(2′),5′-bisphosphate nucleotidase in the NCBI and KEGG databases, that with UniProt ID C4M4T9 is actually a 3′(2′),5′-bisphosphate nucleotidase (Faisal Tarique *et al.*, 2014) and C4M633 in the present study is inositol polyphosphate 1-phosphatase. Hence, these two enzymes are paralogues of each other.

2. Materials and methods

2.1. Preparation of the recombinant vector

The EhIPPase gene (Gene ID 3409888), which was annotated as a 3′(2′),5′-bisphosphate nucleotidase/PAP phosphatase, was PCR-amplified from the genomic DNA of *E. histolytica* HM1IMS using 5′-CATGCCATGGCAATGCAAACTTCGTTATTTGA-3′ (*Nco*I) and 5′-CGGCTCGAGAAGTAAAATTGTTTTTGAAATAGA-3′ (*Xho*I) as the forward and reverse primers, respectively. The amplified product of EhIPPase (858 bp) was digested with *Nco*I and *Xho*I and ligated into pET-28b expression vector with a C-terminal His tag. The absence of any mutation was checked by gene sequencing.

2.2. Expression and purification of EhIPPase in *Escherichia coli*

pET-28b-EhIPPase plasmids were transformed into *E. coli* strain BL21(DE3). The positive colonies were picked from kanamycin-containing LB plates and were inoculated into 50 ml TB medium for overnight growth. 1% of this culture was inoculated into a secondary culture containing 50 mg l^{-1} kanamycin. The culture was incubated at 37°C until the OD_{600} reached 0.6–0.8. Protein expression was initiated with the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 200 μM. The temperature was lowered to 16°C after induction and the culture was allowed to grow overnight. Cells were harvested at 4°C by centrifugation at 8000 rev min^{-1} for 5 min and the pellet was suspended in suspension buffer (at 1 g of pellet per 10 ml) consisting of 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM phenylmethanesulfonyl fluoride (PMSF), 0.1% Triton X-100, 0.2 mM EDTA, 10 mM imidazole, 0.05 mg ml^{-1} lysozyme, 5 mM β-mercaptoethanol (βMe). Lysed cells were sonicated and then centrifuged at 18 000 rev min^{-1} for 1 h. The cleared supernatant was applied onto an Ni Sepharose column and the column was equilibrated with 50 ml washing buffer (50 mM Tris pH 7.5,

150 m*M* NaCl, 5 m*M* βMe, 20 m*M* imidazole). The protein was eluted in buffer consisting of 50 m*M* Tris pH 7.5, 150 m*M* NaCl, 100 m*M* imidazole, 5% glycerol, 5 m*M* βMe. For further purification and the removal of imidazole, the concentrated protein was loaded onto a HiLoad G-200 16/200 column and the flow rate was kept at 1 ml min^{-1}. The purity of the desired protein was checked on 12% SDS–PAGE and the purified fractions were pooled and concentrated using Centricon filter units. The column was equilibrated with running buffer consisting of 10 m*M* Tris pH 7.5, 150 m*M* NaCl, 5% glycerol, 5 m*M* βMe. Glycerol was added to increase the solubility and stability of the protein. The overexpression and purification of selenomethionine-labelled EhIPPase were carried out on the basis of a standardized protocol as described previously (Abdul Rehman *et al.*, 2013).

The identity of the EhIPPase protein was confirmed by tryptic digestion and mass fingerprinting using a Bruker ESI-ion trap mass spectrophotometer. Protein concentration was determined by measuring the UV absorption at 280 nm. An approximate extinction coefficient of 30 830 M^{-1} cm^{-1} was calculated using the *ExPASy* server. PAP phosphatase-1 was expressed and purified as reported previously (Faisal Tarique *et al.*, 2014).

2.3. Enzyme assay

The nucleotidase activity of EhIPPase was measured by the Baykov malachite green method using PAP and Ins(1,4)P2 as substrates (Baykov *et al.*, 1988). For the EhIPPase assay, the standard 100 µl reaction mixture volume consisted of 50 m*M* Tris pH 7.5, 1 m*M* PAP or 1 m*M* Ins(1,4)P2, 5 m*M* $MgCl_2$ and 2 µg purified EhIPPase protein. The reaction mixture was incubated for 5 min at room temperature and then stopped by the addition of 50 µl malachite green dye solution. After 20 min, the blue-coloured phosphomolybdate complex was detected at 650 nm in a microplate reader. The absorbance and hence the activity are directly proportional to the amount of P_i liberated that formed the phosphomolybdate–malachite green complex. To determine the optimum temperature, the optimum pH, the Li^+/Na^+ sensitivity and the nature of the divalent metal cation that acts as a cofactor for the activity of EhIPPase, these factors were varied during the assay. For divalent metal ions, 5 m*M* $MgCl_2$, $CoCl_2$, $MnCl_2$ or $CaCl_2$ were used.

2.4. Dynamic light-scattering measurements

Dynamic light-scattering (DLS) measurements were performed at 830 nm using SpectroSize300 dynamic light-scattering equipment (Molecular Dimensions). The concentration of EhIPPase used for this experiment was around 5 mg ml^{-1}. The protein was centrifuged at 13 000 rev min^{-1} for 30 min at 4°C and filtered through a 0.25 µm Whatman syringe filter prior to the experiment. For each experiment, ten scans were taken; the duration of each scan was 20 s. The mean hydrodynamic radius (Morris *et al.*, 2004), standard deviation, polydispersity and percent of peak area were analyzed using the *DYNAMICS* 6.10 software using the optimized resolution.

2.5. Generation of antibodies and Western blotting

Western blotting was performed by transfer of the protein separated by SDS–PAGE to polyvinylidenedifluoride (PVDF) membrane using a semi-dry transfer system (Bio-Rad). SDS was included in the transfer buffer to increase the transfer efficiency of high-molecular-weight proteins. The membrane was blocked with 5% Blotto (Genotech, USA) in Tris-buffered saline with Tween (TBST) for 1 h. Protein bands were probed with anti-EhIPPase (1:5000) raised in mice, washed with TBST, incubated with a 1:10 000 dilution of HRP-conjugated secondary antibody (Bio-Rad) and finally detected by the enhanced chemiluminescence method.

2.6. Immunofluorescence labelling

E. histolytica HM-1-IMS cells were resuspended in TYI-S-33 incomplete medium and placed on a cover slip wiped with acetone for 10–20 min at 37°C. The medium was decanted and the cells were fixed by keeping them in 3.7% paraformaldehyde (PFA) at 37°C for 30 min. PBS was used to remove PFA and the cells were permeabilized by adding 1 ml 0.1% Triton X-100 for 15 min at 37°C. They were again washed with PBS and neutralized with 50 m*M* NH_4Cl for 30 min at room temperature. Excess NH_4Cl was removed by washing with PBS. The cells were incubated with primary anti-EhIPPase (1:200) at 37°C for 1 h. Prior to the incubation with secondary antibodies, the slides were washed with 1% BSA in PBS. 100 µl DAPI dye was added onto the slide and incubated for 30 min at 37°C. Anti-rabbit Alexa Fluor 594 for EhIPPase was used. The cover slips were sealed and an Olympus Fluoview FV100 laser scanning microscope was used to view confocal images.

2.7. Crystallization and X-ray data collection

The purified EhIPPase protein was concentrated to 8 mg ml^{-1} in 10 m*M* Tris pH 8.0 containing 2 m*M* $MgCl_2$, 150 m*M* NaCl, 1 m*M* ammonium phosphate, 5% glycerol and 5 m*M* βMe. Following extensive robotic crystallization trials (using a Mosquito robot) by hanging-drop vapour diffusion, crystal hits in the form of fine crystalline needles were obtained from condition C4 of the Morpheus screen (Molecular Dimensions). The 500 nl drop consisted of a 1:1 ratio of protein solution and precipitant equilibrated against 200 µl precipitant. Crystals appeared overnight at 16°C. After rigorous optimization, good-quality crystals appeared using a mixture of 12.5% 2-methyl-2,4-pentanediol (MPD), 12.5% PEG 3350 and 12.5% PEG 1000 as precipitants at pH 6.7–6.9 and NaI, NaBr and NaF (as a halogen mixture) at varied concentrations. A single crystal was carefully picked from the clusters, washed and equilibrated with a cryoprotectant solution consisting of 12.5% MPD, 12.5% PEG 3350 and 12.5% PEG 1000 at pH 6.7. These crystals were mounted in cryo-loops, flash-cooled in liquid nitrogen at 100 K and diffracted X-rays to a resolution of 2.54 Å using a Rigaku FR-E+ SuperBright microfocus rotating-anode generator; diffraction images were recorded at various angles on an R-AXIS IV^{++} detector at the National Institute of Immunology (NII), New

Delhi, India and the native data set was indexed, processed and scaled with *HKL*-2000 (Otwinowski & Minor, 1997). For experimental phasing, a selenomethionine (SeMet) single-wavelength anomalous dispersion (SAD) data set was collected from another crystal, which diffracted to a resolution of 2.7 Å on BM14 at ESRF, Grenoble, France (λ = 0.977 Å). The data were indexed, processed and scaled with *HKL*-2000.

2.8. Structure determination and refinement

The structure was solved using the single-wavelength anomalous scattering protocol of *Auto-Rickshaw*, the EMBL-Hamburg automated crystal structure-determination platform (Panjikar *et al.*, 2005). The input diffraction data were prepared and converted for use in *Auto-Rickshaw* using programs from the *CCP*4 suite (Winn *et al.*, 2011). Heavy-atom structure-factor (F_A) values were calculated using *SHELXC* (Sheldrick, 2008). Based on an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 3.2 Å. All of the 12 heavy atoms (Se) in the protein were found using *SHELXD* (Sheldrick, 2008). The correct hand for the substructure was determined using *ABS* (Wang *et al.*, 2004) and *SHELXE* (Sheldrick, 2008). The occupancy of all substructure atoms was refined and the initial phases were calculated using *MLPHARE* from *CCP*4 (Winn *et al.*, 2011). The twofold noncrystallographic symmetry (NCS) operator was found using *RESOLVE* (Terwilliger, 2000). Density modification, phase extension and NCS averaging were performed using *DM* (Cowtan & Main, 1996). A partial α-helical model was produced using *HELICAP* (Morris *et al.*, 2004). The partial model contained 495 residues out of the total of 570 residues, with R and R_{free} values of 24 and 34%. Taking this partial model as a template, the EhIPPase structure was solved by molecular replacement using the higher resolution native data. The model fitted well in the electron-density map and any ambiguous density was manually checked in *Coot* (Emsley & Cowtan, 2004). After multiple rounds of manual building and refinement with *REFMAC* (Murshudov *et al.*, 2011), the values of R and R_{free} decreased to 22 and 25%, respectively. The model at this stage was submitted to the *PDB_REDO* web server for improvement of the refinement statistics, which decreased the R and R_{free} value to 19 and 25 with no ambiguity in the electron density (Joosten *et al.*, 2012). The final model consists of 4574 protein atoms, 33 water molecules, two phosphate ions and six magnesium ions with good refinement statistics (Table 1). The quality of the model was checked using *PROCHECK* (Laskowski *et al.*, 1993). The identities of the metal and PO_4^{3-} ions were confirmed by their binding geometry, coordination sphere, temperature factors and difference Fourier electron-density maps. The coordinates of EhIPPase and its structure factors have been deposited in the Protein Data Bank as PDB entry 4qxd.

2.9. Phylogenetic analysis

Different structures of members of the IMPase superfamily from organisms from the Archaea, Monera, Fungi, Plantae and Animalia were obtained from the PDB and considered for phylogenetic studies. Evolutionary relationships were studied using the *MEGA*4.0 phylogenetic analysis package (Tamura *et al.*, 2007).

Table 1
Data statistics for EhIPPase.

Values in parentheses are for the highest resolution shell.

Data set	EhIPPase	EhIPPase (peak)
Crystallographic data		
X-ray source	NII	BM14, ESRF, France
Wavelength (Å)	1.54	0.977
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 72.7, b = 76.3, c = 101.0, $\alpha = \beta = \gamma$ = 90.0	a = 72.9, b = 77.1, c = 101.0, $\alpha = \beta = \gamma$ = 90.0
Total No. of observations	229209	230607
No. of unique observations	19012	29692
R_{sym} or R_{merge} (%)	10 (50)	9.7 (73)
Completeness (%)	99.9 (99.8)	100 (100)
Multiplicity	12.1 (11.6)	7.8 (7.3)
Average $I/\sigma(I)$	24.3 (4.3)	23.1 (2.7)
No. of molecules in asymmetric unit	2	2
Refinement		
Resolution	50–2.55 (2.64–2.55)	50–2.75 (2.75–2.70)
R_{work}/R_{free} (%)	19/25	24/34
Mean B factor (Å^2)	31.3	
No. of atoms		
Proteins	4574	
Mg	6	
Phosphate	10	
Water	33	
R.m.s. deviations		
Bonds (Å)	0.007	
Bond angles (°)	1.187	
Ramachandran plot (%)		
Most favoured region	88	
Additionally allowed region	11.5	
Generously allowed region	0.6	
Disallowed region	0.0	
PDB code	4qxd	

3. Results

3.1. Studying EhIPPase activity under variable temperature, pH and metal-ion conditions

As the enzyme was annotated as a 3′(2′),5′-bisphosphate nucleotidase, all of the initial kinetic studies were carried out by keeping PAP as the primary substrate. The activity of the EhIPPase was checked with different pH values, temperatures and metal ions. As measured by the ability of the enzyme to hydrolyze PAP in wide ranges of pH, it was found to be active (and hence stable) from pH 7 to 8, with a maximum activity at about pH 7.5 (100%); the enzyme, however, loses its activity under acidic and highly alkaline conditions (Supplementary Fig. S1*a*[1]). By monitoring EhIPPase at different temperatures at pH 7.5, its maximum activity was found to occur at 60°C. However, activity can be observed from 30 to 80°C (Supplementary Fig. S1*b*). All subsequent assays were carried out at pH 7.5 and 30°C unless stated otherwise. With metal ions,

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: RR5078).

EhIPPase showed maximum activity with Mg^{2+}, followed by Co^{2+}, Ca^{2+} and Mn^{2+} (Supplementary Fig. S1*c*). As the activity was found to be maximum with Mg^{2+}, which is also the natural cofactor of this superfamily of phosphatases, it was chosen for carrying out all of the enzymatic studies related to EhIPPase. It was only in the later stages that it was realised that the enzyme was better at hydrolyzing Ins(1,4)P2 than PAP under equivalent conditions. Activity with Ins(1,4)P2 was only checked under the standard conditions of pH 7.5, 5 m*M* $MgCl_2$ and 1 m*M* substrate.

3.2. Localization of EhIPPase

The presence of EhIPPase in the amoebal cell lysate was detected by a Western blotting experiment. The anti-EhIPPase antibody stained endogenous EhIPPase bands at 32 kDa, thus confirming the constitutive expression in *E. histolytica* trophozoites (Supplementary Fig. S2*a*). Through immunofluorescence labeling, it was found that EhIPPase is present throughout the cytoplasm of *E. histolytica* cells (Supplementary Fig. S2*b*).

3.3. Crystal structure of EhIPPase

The crystal structure of native EhIPPase at 2.55 Å resolution has been refined to an R and R_{free} of 19 and 25%, respectively. Fig. 1 shows the quality of the electron density in one of the regions with the best real-space R factors in the structure. Over 88% of the total residues (570 amino acids) were in the most favourable region of the Ramachandran plot, 11.5% of the residues were in the additional allowed region and 0.6% were in the generously allowed region. There are two molecules in the asymmetric unit, which do not interact with each other. This is consistent with the DLS and gel-filtration experiments, in which the active form of EhIPPase was found to be a monomer (Supplementary Figs. S3 and S4).

Similar to other members of the Li^{+}-sensitive, Mg^{2+}-dependent superfamily of phosphatases, the EhIPPase structure is divided into two domains connected to each other by a 12–13-residue loop. The EhIPPase monomer consists of six α-helices, 17 β-strands and six short 3_{10}-helices (Fig. 1 and Supplementary Fig. S5). Residues 1–152 in the N-terminal domain constitute a predominantly $\alpha+\beta$ fold that contains helices H1–H3 and antiparallel β-sheets formed of strands B1–B9. Residues 168–288 of the C-terminal domain form an α/β fold in which strands B11–B17 form mixed parallel/antiparallel β-sheets that are sandwiched between helices H4–H6. The active site of EhIPPase lies at the interface of the domains and was identified by the presence of three metal ions (Mg^{2+}) and one PO_4^{3-} ion in the crystal structure. The identities of the metal ions were confirmed by their binding geometry, coordination sphere, temperature factors and difference Fourier electron-density maps. The structure of EhIPPase bound to metal ions and PO_4^{3-} represents the stage after substrate hydrolysis in which the product of hydrolysis (PO_4^{3-}) was found trapped in the structure.

3.4. Geometry of the active site

As seen in the crystal structure, the active site of EhIPPase has three metal-binding sites named M1, M2 and M3 in a hydrophilic cavity. Amino acids interacting with the metal ions and the substrate/product at the active site and their respective atomic distances are listed in Fig. 2. These residues, which take part in hydrogen-bonding, ionic and hydrophobic interactions with the metal ions and substrates, have been found to be conserved or to be replaced by similar residues in other homologous PAP phosphatases (Albert *et al.*, 2000; Patel *et al.*, 2002).

3.5. Comparison of EhIPPase with PAP phosphatase-1 from *E. histolytica*, the dual-activity PAP phosphatase from rat (RnPIP), PAP phosphatase from yeast (Hal2p) and bovine IPPase

EhIPPase shares little sequence identity with the dual-activity inositol polyphosphate 1-phosphatase/PAP phosphatase from rat (RnPIP), bovine inositol polyphosphate 1-phosphatase (IPPase), PAP phosphatase from yeast (Hal2p) and PAP phosphatase-1 from *E. histolytica*. Despite this, all of them share similar folds and the basic core structure for metal binding and catalysis. Pairwise and three-dimensional structural superposition of EhIPPase with homologous and heterologous structures was performed to analyze the structural differences (Table 2). All of them share conserved signature sequence motifs for metal binding, *i.e.* D-X_n-EE-X_n-DP(I/L)DG(S/T)-X_n-WD-X_n-GG, which is a hallmark feature

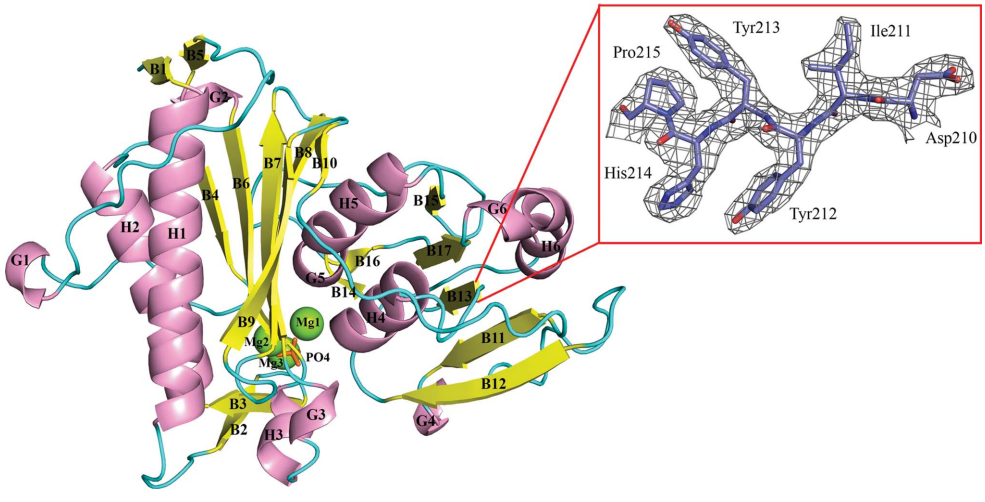

Figure 1
(*a*) Ribbon representations of EhIPPase with bound metal ions and PO_4^{3-} in the active site. (*b*) $2F_o - F_c$ electron density at a 1σ cutoff for a portion of the B13 strand.

of the members of the inositol monophosphatase superfamily of proteins (Supplementary Fig. S6; York *et al.*, 1995). There was a significant degree of divergence owing to the orientations and types of the secondary-structure elements between the EhIPPase, RnPIP and bovine IPPase structures (Fig. 3).

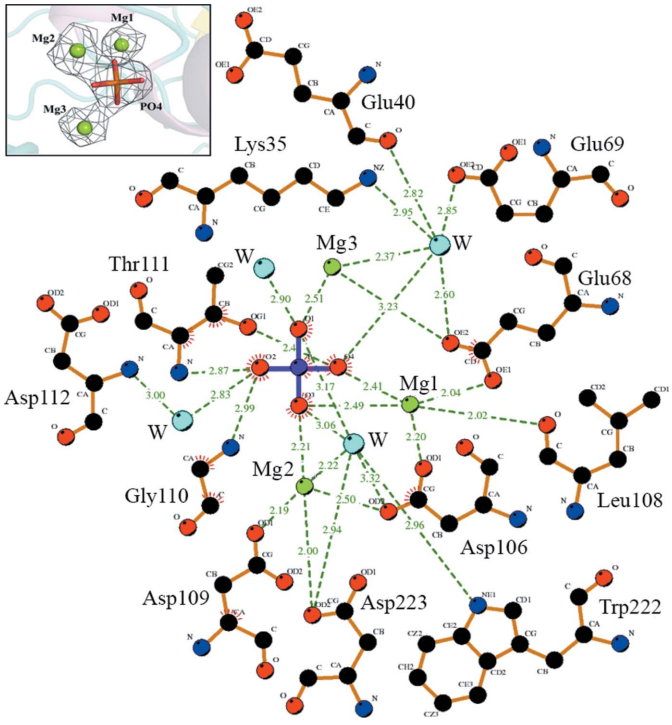

Figure 2
Metal-ion interactions with the ligands. The interactions of the EhIPPase protein with bound Mg^{2+} metal ions and PO_4^{3-} at the active site. The plots were generated by *LigPlot+* (Laskowski & Swindells, 2011). Hydrogen bonds are shown as green dotted lines, while arcs represent residues that make nonbonded contacts with the ligands. The bound metal ions are shown in green and waters in light blue.

Table 2
Structural superposition of members of the Li^{+}-sensitive/Mg^{2+}-dependent phosphatase superfamily with EhIPPase.

The coordinates used for structural comparison and phylogenetic studies are as follows: 4o7i, PAP phosphatase-1 (*E. histolytica*); 1jp4, inositol polyphosphate 1-phosphatase (IPPase)/PAP phosphatase (rat); 1k9y, PAP phosphatase (yeast); 2hhm, IMPase (human); 2wef, PAP phosphatase (human); 1inp, IPPase (bovine); 1lbx, FBPase/IMPase (*A. fulgidus*); 1frp, FBPase (pig); 1dk4, IMPase/FBPase (*M. jannaschii*); 1ime, IMPase (human); 2bji, IMPase (*Bos taurus*); 4as5, IMPase (rat); 2czh, IMPase (human); 2fvz, IMPase (human); 2q74, Suh B (*Mycobacterium tuberculosis*); 3t0j, IMPase (*S. aureus*); 4g60, IMPase/NADP phosphatase (*S. aureus*); 3qmf, IMPase (*S. aureus*); 2p3n, IPPase (*Thermotoga maritima*); 2qfl, IMPase (*E. coli*); 2pcr, IMPase (*Aquifex aeolicus*); 1k9y, PAP phosphatase (yeast); 4j13, EhIPPase (*E. histolytica*); 1vdw, IMPase (*Pyrococcus horikoshii*); 1g0i, FBPase/IMPase (*M. jannaschii*); 1fpi, FBPase (*Sus scrofa*).

Structural alignment of EhIPPase with	Sequence identity (%)	No. of C^{α} atoms aligned	R.m.s.d. (Å)
4o7i/4hxv	22	258	3.025
1jp4 (Patel *et al.*, 2002)	27	282	2.620
1k9y (Albert *et al.*, 2000)	23	256	3.011
2hhm (Bone *et al.*, 1992)	22	244	3.219
2wef (Structural Genomics Consortium, unpublished work)	27	281	2.625
1inp (York *et al.*, 1994)	24	274	3.057
1lbx (Stieglitz *et al.*, 2002)	16	237	3.032
1frp (Xue *et al.*, 1994)	18	227	3.025
1dk4 (Stec *et al.*, 2000)	17	233	3.102

3.6. EhIPPase is specific for Ins(1,4)P2, while PAP phosphatase-1 is specific for PAP

The incubation of equal amounts of PAP phosphatase-1 and EhIPPase with PAP and Ins(1,4)P2 under equivalent reaction conditions leads to a difference in the amount of free PO_4^{3-} liberated from their cognate substrates. The experiments showed that PAP phosphatase-1 is maximally active with PAP as the substrate but showed less than 20% activity with Ins(1,4)P2. In contrast, EhIPPase showed maximum activity

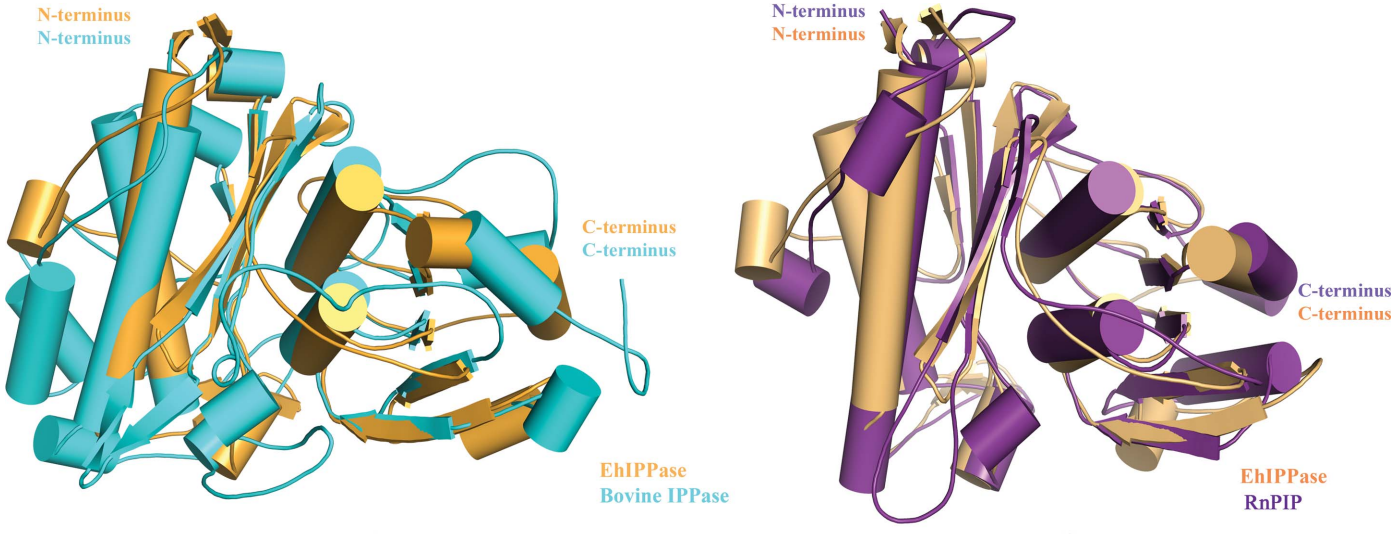

Figure 3
Structural superposition. Structural superpositions of EhIPPase (gold) on (*a*) bovine IPPase (cyan) and (*b*) RnPIP (purple) reveal similar core structures but prominent structural differences in terms of the number of secondary-structure elements and their orientations. *FATCAT*, a web-based server, was used for all structural superpositions (Ye & Godzik, 2004).

with Ins(1,4)P2 and less than 10% activity with PAP as the substrate (Fig. 4). [These experiments were compared based on assigning a value of 100% to the activity of EhIPPase with Ins(1,4)P2.] These results show that PAP phosphatase-1 behaves like a 3′(2′),5′-bisphosphate nucleotidase, while EhIPPase behaves like a inositol polyphosphate 1-phosphatase. Nonetheless, both enzymes also showed little activity with their secondary substrates.

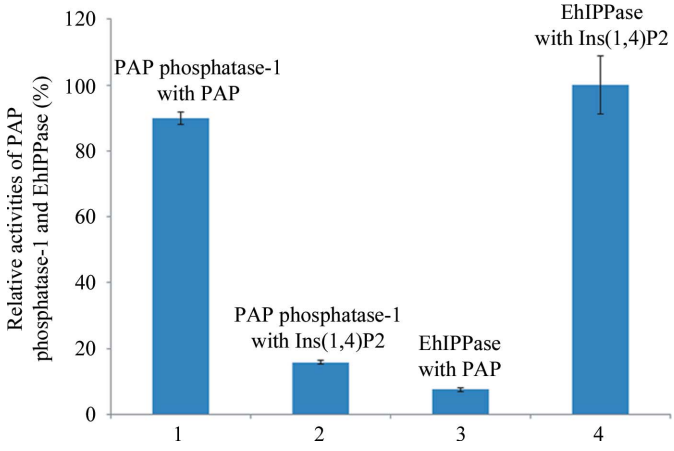

Figure 4
EhIPPase and PAP phosphatase-1 activities with PAP and Ins(1,4) P2 as substrates. Lane 1 shows PAP phosphatase-1 with PAP, lane 2 shows PAP phosphatase-1 with Ins(1,4)P2, lane 3 shows EhIPPase with PAP and lane 4 shows EhIPPase with Ins(1,4)P2. PAP phosphatase-1 hydrolyzes PAP better than does EhIPPase, while EhIPPase hydrolyzes Ins(1,4)P2 better than does PAP phosphatase-1. The standard 100 µl reaction-mixture volume contained 50 m*M* Tris pH 7.5, 5 m*M* $MgCl_2$, 2 µg purified EhIPPase or PAP phosphatase-1 protein and 1 m*M* PAP or Ins(1,4)P2.

The difference in the substrate specificity between PAP phosphatase-1 and EhIPPase can be explained on the basis of their active-site geometries. Close analysis and comparison of PAP phosphatase-1 (*E. histolytica*), EhIPPase, RnPIP [rat PIPase with both IPPase and 3′(2′),5′-bisphosphate nucleotidase activity], bovine IPPase and Hal2p (yeast) structures led us to hypothesize that a couple of secondary structures and adjacent loops (red) modulate the solvent-accessible volume near the active-site pocket, which seems to be responsible for their different substrate specificities (Fig. 5). In PAP phosphatases, *e.g.* Hal2p, RnPIP and PAP phosphatase-1 from *E. histolytica*, these loops have conserved aromatic amino-acid residues for stacking the adenine ring of the PAP substrate and positioning its 3′-phosphate towards the metal-binding sites for efficient hydrolysis (Patel *et al.*, 2002; Albert *et al.*, 2000). In contrast, in EhIPPase the loops (red) lack any aromatic amino-acid binding pocket and have Gln218 and Lys175 protruding into the active-site pocket, which would sterically hinder PAP accessibility, making it difficult to hold the adenine ring of PAP in place for subsequent hydrolysis. These structural features are also seen in bovine IPPase (York

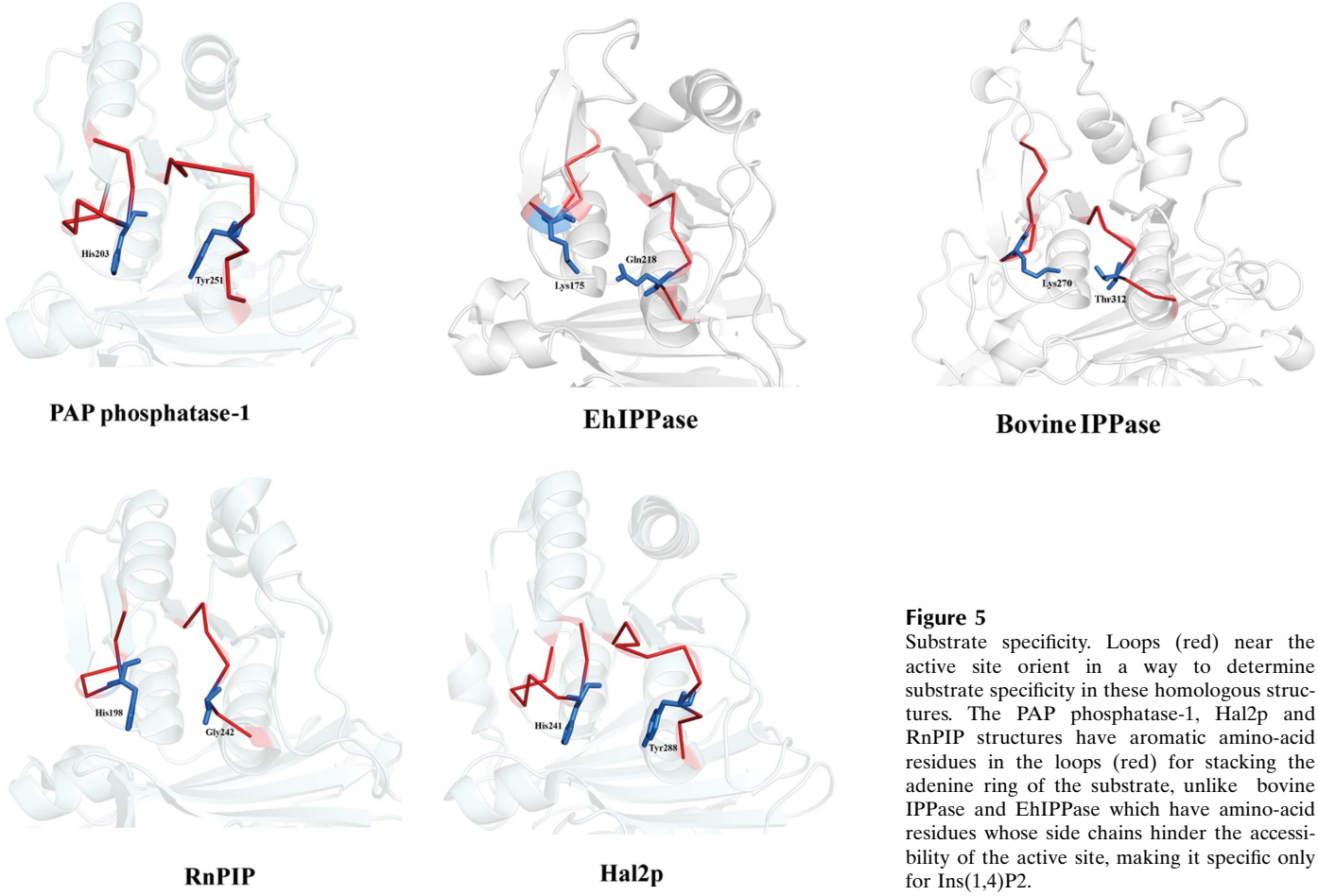

Figure 5
Substrate specificity. Loops (red) near the active site orient in a way to determine substrate specificity in these homologous structures. The PAP phosphatase-1, Hal2p and RnPIP structures have aromatic amino-acid residues in the loops (red) for stacking the adenine ring of the substrate, unlike bovine IPPase and EhIPPase which have amino-acid residues whose side chains hinder the accessibility of the active site, making it specific only for Ins(1,4)P2.

et al., 1994), in which the same type of steric hindrance of PAP is observed owing to Thr312 and Lys270 at equivalent positions, which allowed this IPPase enzyme to react efficiently with Ins(1,4)P2.

Docking studies are also informative in this regard. Molecular docking of Ins(1,4)P2 into the active site of EhIPPase produces proper hydrogen-bond interactions with the 4′-phosphate group of Ins(1,4)P2. The 1′-phosphate group is also found to be coordinated by all three magnesium ions (Fig. 6*a*), holding the substrate in a high-energy state for proper catalysis. These types of arrangements explain why Ins(1,4)P2 might be the natural substrate of EhIPPase. The capacity, albeit at low levels, of EhIPPase to hydrolyze PAP according to kinetic studies can also be explained by molecular docking of AMP (a product of PAP hydrolysis) into the active site of EhIPPase. Docking studies based on the RnPIP structure produced a satisfactory hydrogen-bond network between the framework amino-acid residues at the active site and the sugar and 5′-phosphate moiety of AMP, indicating that the enzyme could also possibly hold PAP (albeit less efficiently), leading to its hydrolysis (Fig. 6*b*). The 4′-phosphate group of Ins(1,4)P2 and the 5′-phosphate group of AMP exist in a similar milieu of amino-acid residues, suggesting a similar expected mode of catalysis for these respective substrates.

Phylogenetic studies have revealed that members of the IMPase superfamily from various organisms are distributed all over the tree; however, a distinct single clade for PAP phosphatases and IPPase can be seen. It is evident from the tree that EhIPPase from *E. histolytica* is phylogenetically different from other typical PAP phosphatases and is more similar to IPPases (Supplementary Fig. S7). The outcomes of the biochemical and phylogenetic studies were found to be consistent with the crystal structure of EhIPPase. The results obtained so far have clearly shown that EhIPPase is capable of hydrolyzing both PAP and Ins(1,4)P2, with the latter acting as its natural substrate since it showed maximum activity.

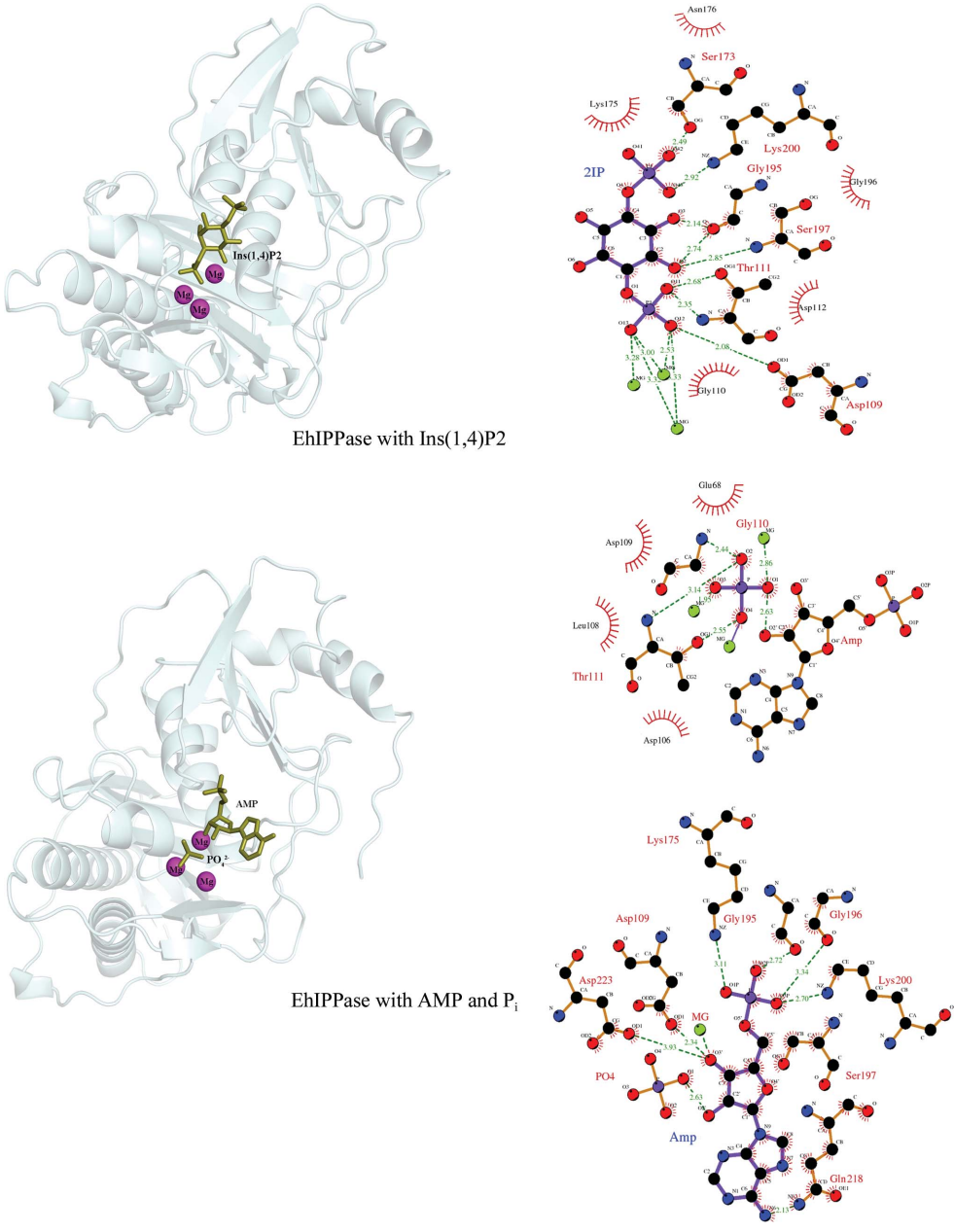

Figure 6
Molecular docking of Ins(1,4)P2 and AMP into the active site of EhIPPase

3.7. Mechanism of action

Comparison of EhIPPase with other structures has provided insight into its probable mechanism of catalysis. The conservation of the DP(I/L)D(G/S)T residues that are involved in catalysis by RnPiP and bovine IPPase extend to EhIPPase (Fig. 7). Structural superposition of this conserved sequence motif in RnPIP, bovine IPPase and EhIPPase showed an r.m.s.d. of 0.16, displaying a high degree of structural conservation of these residues in the active site (Ilinkin *et al.*, 2010). Therefore, the mechanism of EhIPPase is expected to be similar to these homologous enzymes (Patel *et al.*, 2002; York *et al.*, 1994), for which a three metal ion-dependent mechanism has been proposed (Supplementary Fig. S8).

Table 3
Approximate IC_{50} (m*M*) of lithium inhibition for members of the IMPase superfamily.

EhIPPase	40
PAP phosphatase-1 (Faisal Tarique *et al.*, 2014)	1
Hal2p (Murguía *et al.*, 1995)	0.07
CysQ from *M. tuberculosis* (Hatzios *et al.*, 2008)	0.5
Human IMPase (Yenush *et al.*, 2000)	0.3
Sal1 from *Arabidopsis* (Quintero *et al.*, 1996)	0.2
Bovine IPPase (Inhorn & Majerus, 1988)	0.3
IMPase from *M. jannaschii* (MJ0109) (Stec *et al.*, 2000)	250
IMPase/FBPase from *A. fulgidus* (Stieglitz *et al.*, 2002)	290
DHAL2 from *Debaryomyces hansenii* (Aggarwal *et al.*, 2005)	2.4
CysQ from *E. coli* (Fukuda *et al.*, 2007)	0.05
Porcine liver FBPase (Zhang *et al.*, 1996)	1–3
RnPIP (IPPase/PAPase) (Patel *et al.*, 2002)	0.8

3.8. Sensitivities towards Li^+ and Na^+ ions

All members of the IMPase superfamily of phosphatases, to which EhIPPase belongs, show some universal features such as inhibition by Li^+ (Li^+ is an uncompetitive inhibitor for most of the members of this superfamily; Bone *et al.*, 1992) and the requirement for divalent metal ions (Mg^{2+}) as a cofactor for catalysis. The sensitivity of EhIPPase towards Li^+ and Na^+ ions was checked in the presence of PAP. Although a few homologous enzymes have been shown to be sensitive towards Na^+, EhIPPase, similar to PAP phosphatase-1, was less sensitive to Na^+ [IC_{50}(NaCl) = 1 *M*; Supplementary Fig. S9]. Compared with other members of its superfamily (Table 3), EhIPPase is moderately sensitive to Li^+ ions, with an IC_{50} for LiCl of 40 m*M* (Supplementary Fig. S9).

Li^+, being small and having only two electrons, is a poor scatterer of X-rays. Crystal structures of various enzymes from the members of the Li^+-sensitive and Mg^{2+}-sensitive IMPase superfamily have been reported, but none of them gave a clear view of the mode of Li^+ inhibition in this class of enzymes. Therefore, any proposed mechanism for Li^+ inhibition of EhIPPase has only been explained on the basis of indirect evidence and previous studies performed on homologous structures. Recent studies of *Staphylococcus aureus* IMPase (Bhattacharyya *et al.*, 2012), bovine IMPase (Gill *et al.*, 2005) and *Methanococcus jannaschii* IMPase (Johnson *et al.*, 2001) have revealed that the affinity of Mg^{2+} for the metal-binding sites decreases from M1 to M3. While the binding of Mg^{2+} at M1 and M2 is shown to be cooperative in nature, it is uncooperative for M3. Owing to the similar ionic radii of Li^+ (0.060 nm) and Mg^{2+} (0.065 nm), there is competition between these ions for these metal-binding sites (Bhattacharyya *et al.*, 2012). However, owing to the stronger affinity of Mg^{2+} for M1 and M2, it is difficult for Li^+ to competitively replace them. On the other hand, Mg^{2+} is a weak binder of M3 and can be replaced by Li^+ at this site (Bhattacharyya *et al.*, 2012).

The replacement of Mg^{2+} by Li^+ at M3 results in uncompetitive inhibition of the enzyme, with the substrate/product trapped inside the active site. The propensity of binding and localization of Li^+/Mg^{2+} at M3 is assisted by a mobile catalytic loop (Stieglitz *et al.*, 2002). It is the conformation of the loop (conserved throughout the IMPase superfamily) near the active site that makes the enzyme sensitive or insensitive towards Li^+ (Bhattacharyya *et al.*, 2012; Gill *et al.*, 2005;

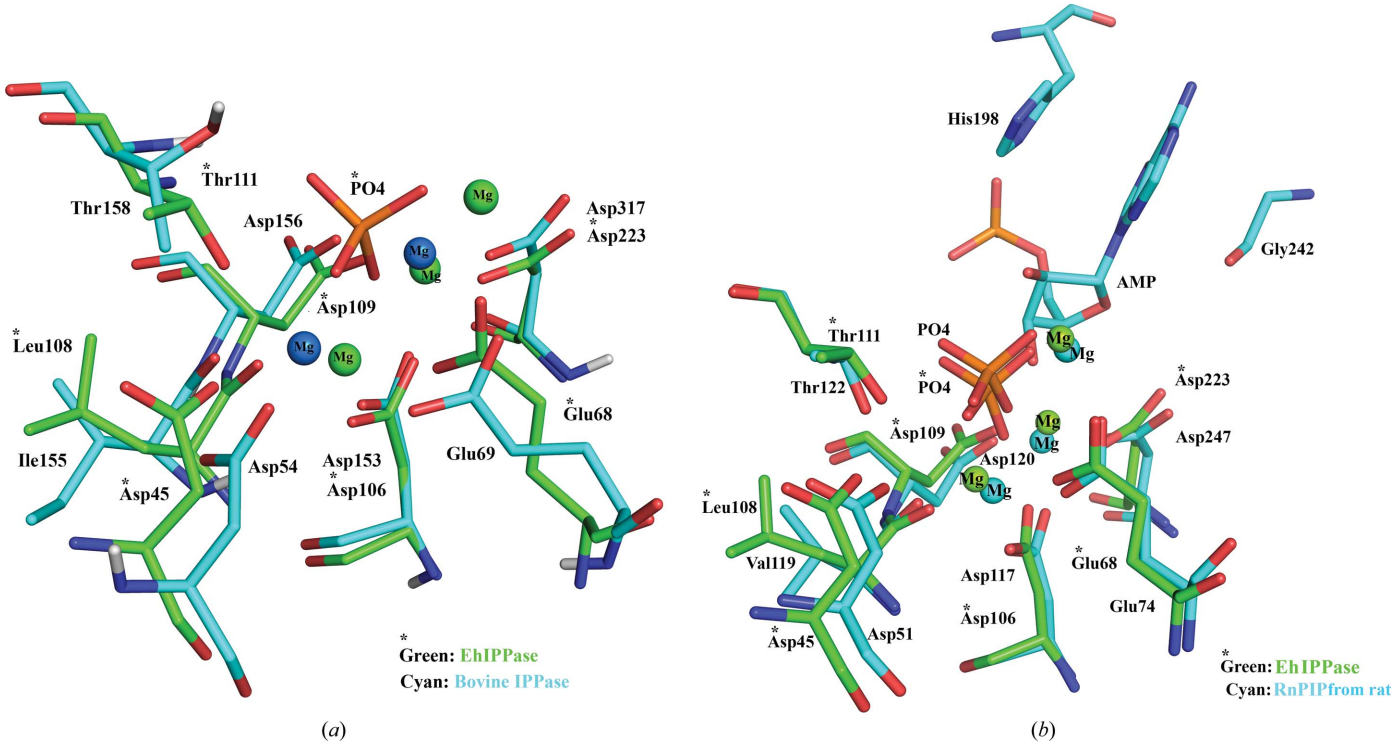

Figure 7
Active-site comparisons. Superposition of the active site of EhIPPase (green) with (*a*) bovine IPPase and (*b*) RnPIP from rat (cyan) shows that the amino-acid residue coordination and the specific orientations of the residues required for the binding of the metal ions are conserved and superpose well with each other.

Stieglitz *et al.*, 2002; Faisal Tarique *et al.*, 2014). It has been observed that if the loop is short, relatively immobile and situated near the active site, the affinity for Mg^{2+} of M3 increases and Mg^{2+} cannot be easily replaced by Li^{+}, making the enzyme insensitive to Li^{+} inhibition (Stieglitz *et al.*, 2002). However, in enzymes that are Li^{+}-sensitive, the catalytic loop is long, mobile and tilted away from the active site. This decreases the affinity of M3 for Mg^{2+}, which eventually becomes replaced by Li^{+} (Stieglitz *et al.*, 2002). The Li^{+} ion, possessing a lower charge, cannot generate the necessary nucleophilic water molecule essential for catalysis, thereby inhibiting the enzyme uncompetitively (Stieglitz *et al.*, 2002).

EhIPPase was superimposed with two Li^{+}-insensitive enzymes, *Archaeoglobus fulgidus* IMPase and *M. jannaschii* IMPase, and with two Li^{+}-sensitive enzymes, yeast PAP phosphatase (Hal2P) and RnPIP, and the conformation of the mobile loop was compared (Supplementary Fig. S10). It was found that the loop of EhIPPase is tilted away from the active site and more closely resembles those of RnPIP and Hal2P, suggesting EhIPPase to be a Li^{+}-sensitive enzyme. This *in silico* result is consistent with the biochemical experiment [$IC_{50}(Li^{+})$ = 40 m*M*] described above. The overall outcome of this work further strengthens the hypothesis regarding the conformation of a mobile catalytic loop determining the sensitivity towards Li^{+} ion. The hairpin flap consisting of residues Lys33–Val41 (Supplementary Fig. S10) is most probably serving the purpose of a catalytic mobile loop in EhIPPase similar to those in the homologous phosphatases mentioned above. EhIPPase showed lower or no sensitivity towards Na^{+}; the hypothesis for Na^{+} inhibition has previously been explained for Hal2p (Albert *et al.*, 2000). It was proposed that a decrease in the physical size of metal-binding site 2 prevents the larger Na^{+} from binding and inhibiting the enzyme (Albert *et al.*, 2000). The same hypothesis can be made in the case of EhIPPase with little modification, in which metal-binding site 3, being the weakest metal-binding site and showing less affinity (assumed to be the binding site of Li^{+}) for metal ions, may be preventing the larger Na^{+} from entering into M3, thereby making it less sensitive towards inhibition.

4. Discussion

Analysis of the *E. histolytica* genome suggested two putative 3′(2′),5′-bisphosphate nucleotidases/PAP phosphatases, NCBI Gene IDs 3406246 and 3409888, which share less than 25% sequence identity. On the basis of the current structural and biochemical studies, NCBI Gene ID 3409888 has been identified as an inositol polyphosphate 1-phosphatase. Comparison of this protein with its PAP phosphatase-1 paralogue (NCBI GeneID 3406246; Faisal Tarique *et al.*, 2014) and other structures of the IMPase superfamily have provided insights into the molecular and functional details of EhIPPase.

Unlike PAP phosphatase-1, which showed maximum activity with and a preference for PAP as the substrate, EhIPPase showed maximum activity with and a preference for Ins(1,4)P2. We thus propose that inside the cell PAP phosphatase-1 is involved primarily in regulating the sulfate-activation pathway by acting as a 3′(2′),5′-bisphosphate nucleotidase, while EhIPPase could be fundamentally involved in the phosphatidyl signalling pathway and lipid metabolism by acting as an inositol polyphosphate 1-phosphatase. The abilities of PAP phosphatase-1 and EhIPPase to hydrolyze their secondary substrates, albeit at lower activity levels, indicate a possible connection between sulfur and inositol 1,4-bisphosphate metabolism which needs to be explored further. RnPIP from rat and the product of the SAL1 gene of *Arabidopsis* are other homologous IPPases which have dual specificity, acting primarily on PAP and with auxiliary action on Ins(1,4)P2 (López-Coronado *et al.*, 1999; Quintero *et al.*, 1996).

Structural comparison with previously solved homologous structures led to the identification of loops near the active site which contain important amino-acid residues necessary for differential binding and hydrolysis of substrates. Based on the active-site geometry and the ability of this enzyme to hydrolyze Ins(1,4)P2 with a much greater preference than for PAP, we conclude that this enzyme is an inositol polyphosphate 1-phosphatase. Superposition of the active-site residues with homologous structures reveals similar spatial orientations of the amino-acid residues responsible for metal binding and substrate hydrolysis, and suggests a similar three metal-dependent catalytic mechanism assisted by a mobile catalytic loop. Based on the conformation of this loop, we propose EhIPPase to be Li^{+}-sensitive, and it has been shown biochemically to be moderately sensitive to this alkali-metal ion. The biochemical, structural, docking and phylogenetic studies are consistent with and complement each other. We conclude that the 'missing' inositol polyphosphate 1-phosphatase of *E. histolytica* is EhIPPase.

5. Related literature

The following references are cited in the Supporting Information for this article: Gouet *et al.* (2003) and McWilliam *et al.* (2013).

We thank the Department of Biotechnology, Government of India for funding. We thank the ESRF, BM14 staff, Department of Biotechnology (DBT), Government of India for access to the beamline and National Institute of Immunology for data collection. KFT and SAAR thank the CSIR for fellowship. We also thank Professor Alok Bhattacharya, Jawaharlal Nehru University for support and providing genomic DNA and *E. histolytica* cells. We thank the Advanced Instrumentation Research Facility (AIRF), Jawaharlal Nehru University (JNU) for crystallization and preliminary data collection. We thank UGC-RNW, DST-PURSE, DST-FIST for funding institutional and central instrumental facility support. CB thanks BMBF for the grant of 01DQ12013.

References

Abdul Rehman, S. A., Verma, V., Mazumder, M., Dhar, S. K. & Gourinath, S. (2013). *J. Bacteriol.* **195**, 2826–2838.

Aggarwal, M., Bansal, P. K. & Mondal, A. K. (2005). *Yeast*, **22**, 457–470.

Albert, A., Yenush, L., Gil-Mascarell, M. R., Rodriguez, P. L., Patel, S., Martínez-Ripoll, M., Blundell, T. L. & Serrano, R. (2000). *J. Mol. Biol.* **295**, 927–938.

Aslam, S., Bhattacharya, S. & Bhattacharya, A. (2012). *PLoS Pathog.* **8**, e1003055.

Baykov, A. A., Evtushenko, O. A. & Avaeva, S. M. (1988). *Anal. Biochem.* **171**, 266–270.

Berridge, M. J. (1984). *Biochem. J.* **220**, 345–360.

Bhattacharya, A., Padhan, N., Jain, R. & Bhattacharya, S. (2006). *Arch. Med. Res.* **37**, 221–225.

Bhattacharyya, S., Dutta, D., Saha, B., Ghosh, A. K. & Das, A. K. (2012). *Biochimie*, **94**, 879–890.

Bone, R., Springer, J. P. & Atack, J. R. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 10031–10035.

Brown, D. A. & London, E. (1998). *Annu. Rev. Cell Dev. Biol.* **14**, 111–136.

Choudhuri, G. & Rangan, M. (2012). *Indian J. Gastroenterol.* **31**, 153–162.

Christy, N. C. & Petri, W. A. Jr (2011). *Future Microbiol.* **6**, 1501–1519.

Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Faisal Tarique, K., Arif Abdul Rehman, S. & Gourinath, S. (2014). *Acta Cryst.* D**70**, 2019–2031.

Fukuda, C., Kawai, S. & Murata, K. (2007). *Appl. Environ. Microbiol.* **73**, 5447–5452.

Gamper, N. & Shapiro, M. S. (2007). *J. Physiol.* **582**, 967–975.

Gilchrist, C. A. & Petri, W. A. (1999). *Curr. Opin. Microbiol.* **2**, 433–437.

Gill, R., Mohammed, F., Badyal, R., Coates, L., Erskine, P., Thompson, D., Cooper, J., Gore, M. & Wood, S. (2005). *Acta Cryst.* D**61**, 545–555.

Goldston, A. M., Powell, R. R., Koushik, A. B. & Temesvari, L. A. (2012). *Eukaryot. Cell*, **11**, 743–751.

Goldston, A. M., Powell, R. R. & Temesvari, L. A. (2012). *Trends Parasitol.* **28**, 417–426.

Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.

Grewal, J. S., Padhan, N., Aslam, S., Bhattacharya, A. & Lohia, A. (2013). *Cell. Microbiol.* **15**, 2020–2033.

Hatzios, S. K., Iavarone, A. T. & Bertozzi, C. R. (2008). *Biochemistry*, **47**, 5823–5831.

Ilinkin, I., Ye, J. & Janardan, R. (2010). *BMC Bioinformatics*, **11**, 71.

Inhorn, R. C. & Majerus, P. W. (1988). *J. Biol. Chem.* **263**, 14559–14565.

Johnson, K. A., Chen, L., Yang, H., Roberts, M. F. & Stec, B. (2001). *Biochemistry*, **40**, 618–630.

Joosten, R. P., Joosten, K., Murshudov, G. N. & Perrakis, A. (2012). *Acta Cryst.* D**68**, 484–496.

Koushik, A. B., Powell, R. R. & Temesvari, L. A. (2013). *Infect. Immun.* **81**, 2145–2155.

Koushik, A. B., Welter, B. H., Rock, M. L. & Temesvari, L. A. (2014). *Eukaryot. Cell*, **13**, 401–411.

Kumar, S., Padhan, N., Alam, N. & Gourinath, S. (2007). *Proteins*, **68**, 990–998.

Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.

Laskowski, R. A. & Swindells, M. B. (2011). *J. Chem. Inf. Model.* **51**, 2778–2786.

Laughlin, R. C., McGugan, G. C., Powell, R. R., Welter, B. H. & Temesvari, L. A. (2004). *Infect. Immun.* **72**, 5349–5357.

Lemmon, M. A. (2008). *Nature Rev. Mol. Cell Biol.* **9**, 99–111.

López-Contreras, L., Hernández-Ramírez, V. I., Flores-García, Y., Chávez-Munguía, B. & Talamás-Rohana, P. (2013). *Parasitology*, **140**, 202–209.

López-Coronado, J. M., Bellés, J. M., Lesage, F., Serrano, R. & Rodríguez, P. L. (1999). *J. Biol. Chem.* **274**, 16034–16039.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P. & Lopez, R. (2013). *Nucleic Acids Res.* **41**, W597–W600.

Morris, R. J., Zwart, P. H., Cohen, S., Fernandez, F. J., Kakaris, M., Kirillova, O., Vonrhein, C., Perrakis, A. & Lamzin, V. S. (2004). *J. Synchrotron Rad.* **11**, 56–59.

Murguía, J. R., Bellés, J. M. & Serrano, R. (1995). *Science*, **267**, 232–234.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pacheco, J., Shibayama, M., Campos, R., Beck, D. L., Houpt, E., Petri, W. A. Jr & Tsutsumi, V. (2004). *Parasitol. Int.* **53**, 35–47.

Padilla-Vaca, F. & Anaya-Velázquez, F. (2010). *Infect. Disord. Drug Targets*, **10**, 242–250.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* D**61**, 449–457.

Patel, S., Yenush, L., Rodríguez, P. L., Serrano, R. & Blundell, T. L. (2002). *J. Mol. Biol.* **315**, 677–685.

Powell, R. R., Welter, B. H., Hwu, R., Bowersox, B., Attaway, C. & Temesvari, L. A. (2006). *Exp. Parasitol.* **112**, 221–231.

Quintero, F. J., Garciadeblás, B. & Rodríguez-Navarro, A. (1996). *Plant Cell*, **8**, 529–537.

Rivière, C., Marion, S., Guillén, N., Bacri, J.-C., Gazeau, F. & Wilhelm, C. (2007). *J. Biomech.* **40**, 64–77.

Salles, J. M., Salles, M. J., Moraes, L. A. & Silva, M. C. (2007). *Expert Rev. Anti Infect. Ther.* **5**, 893–901.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Stec, B., Yang, H., Johnson, K. A., Chen, L. & Roberts, M. F. (2000). *Nature Struct. Biol.* **7**, 1046–1050.

Stieglitz, K. A., Johnson, K. A., Yang, H., Roberts, M. F., Seaton, B. A., Head, J. F. & Stec, B. (2002). *J. Biol. Chem.* **277**, 22863–22874.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). *Mol. Biol. Evol.* **24**, 1596–1599.

Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.

Wang, J. W., Chen, J. R., Gu, Y. X., Zheng, C. D., Jiang, F., Fan, H. F., Terwilliger, T. C. & Hao, Q. (2004). *Acta Cryst.* D**60**, 1244–1253.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Xue, Y., Huang, S., Liang, J.-Y., Zhang, Y. & Lipscomb, W. N. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 12482–12486.

Ye, Y. & Godzik, A. (2004). *Nucleic Acids Res.* **32**, W582–W585.

Yenush, L., Bellés, J. M., López-Coronado, J. M., Gil-Mascarell, R., Serrano, R. & Rodríguez, P. L. (2000). *FEBS Lett.* **467**, 321–325.

York, J. D., Ponder, J. W., Chen, Z., Mathews, F. S. & Majerus, P. W. (1994). *Biochemistry*, **33**, 13164–13171.

York, J. D., Ponder, J. W. & Majerus, P. W. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 5149–5153.

Zhang, R., Villeret, V., Lipscomb, W. N. & Fromm, H. J. (1996). *Biochemistry*, **35**, 3038–3043.