BIOLOGICAL CRYSTALLOGRAPHY
ISSN 1399-0047

Cap-domain closure enables diverse substrate recognition by the C2-type haloacid dehalogenase-like sugar phosphatase *Plasmodium falciparum* HAD1

Jooyoung Park,[a] Ann M. Guggisberg,[b] Audrey R. Odom[a,b] and Niraj H. Tolia[a,c]*

[a]Department of Molecular Microbiology, Washington University School of Medicine, St Louis, MO 63110, USA, [b]Department of Pediatrics, Washington University School of Medicine, St Louis, MO 63110, USA, and [c]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St Louis, MO 63110, USA. *Correspondence e-mail: tolia@wustl.edu

Received 20 April 2015
Accepted 23 June 2015

Edited by R. McKenna, University of Florida, USA

Keywords: haloacid dehalogenase; HAD superfamily; C2 cap; substrate specificity; sugar phosphatase.

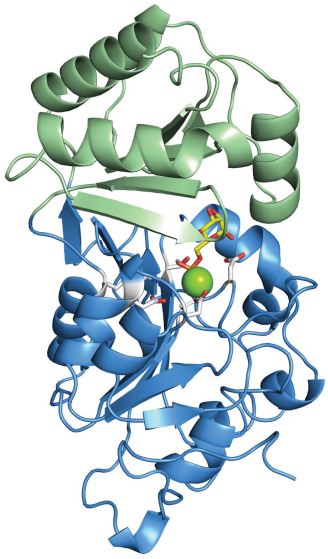
PDB references: PfHAD1 + Man6P, 4zev; PfHAD1 + Glu6P, 4zew; PfHAD1 + Gly3P, 4zex

Supporting information: this article has supporting information at journals.iucr.org/d

Haloacid dehalogenases (HADs) are a large enzyme superfamily of more than 500 000 members with roles in numerous metabolic pathways. *Plasmodium falciparum* HAD1 (PfHAD1) is a sugar phosphatase that regulates the methylerythritol phosphate (MEP) pathway for isoprenoid synthesis in malaria parasites. However, the structural determinants for diverse substrate recognition by HADs are unknown. Here, crystal structures were determined of PfHAD1 in complex with three sugar phosphates selected from a panel of diverse substrates that it utilizes. Cap-open and cap-closed conformations are observed, with cap closure facilitating substrate binding and ordering. These structural changes define the role of cap movement within the major subcategory of C2 HAD enzymes. The structures of an HAD bound to multiple substrates identifies binding and specificity-determining residues that define the structural basis for substrate recognition and catalysis within the HAD superfamily. While the substrate-binding region of the cap domain is flexible in the open conformations, this region becomes ordered and makes direct interactions with the substrate in the closed conformations. These studies further inform the structural and biochemical basis for catalysis within a large superfamily of HAD enzymes with diverse functions.

1. Introduction

The HAD superfamily, the name of which derives from bacterial haloacid dehalogenases, is one of the largest classes of enzymes and is represented throughout all kingdoms of life (Koonin & Tatusov, 1994). This superfamily is comprised of more than 500 000 members (InterPro IPR023214; Hunter *et al.*, 2012) that have evolved to serve diverse biological functions. HADs have been shown to play roles in primary (Collet *et al.*, 1997; Rangarajan *et al.*, 2006) and secondary metabolism (Wang *et al.*, 2008; Lu *et al.*, 2009; Wu & Woodard, 2003; Biswas *et al.*, 2009), regulation of metabolic pools (Rinaldo-Matthis *et al.*, 2002), cell housekeeping (Kim *et al.*, 2004; Fortpied *et al.*, 2006; Kuznetsova *et al.*, 2006; Titz *et al.*, 2007) and nutrient uptake (Passariello *et al.*, 2006). HADs are promiscuous enzymes, displaying catalytic activity towards a broad range of substrates (Pandya *et al.*, 2014). The majority of HADs are involved in variations of phosphoryl-transfer reactions (Burroughs *et al.*, 2006) and include phosphoesterases, ATPases, phosphonatases and sugar phosphomutases (Allen & Dunaway-Mariano, 2004, 2009). HADs contain four highly conserved sequence motifs, which coordinate a magnesium ion and bind the phosphoryl group of



substrate compounds. Together, these HAD-motif residues comprise a self-contained scaffold for catalysis through a phosphoaspartyl intermediate (Allen & Dunaway-Mariano, 2009; Seifried *et al.*, 2013; Collet *et al.*, 2002). Further biochemical and structural studies, combined with biological function studies, are needed to understand how substrate recognition and catalysis are achieved by these promiscuous enzymes in order to carry out their diverse metabolic regulatory functions.

We recently identified and reported the crystal structure of the first HAD protein from *Plasmodium falciparum*, which we called PfHAD1 (Guggisberg *et al.*, 2014). PfHAD1 is a sugar phosphatase that regulates substrate availability to the methylerythritol phosphate (MEP) pathway for isoprenoid precursor biosynthesis (Fig. 1). Understanding the regulation of the MEP pathway is important, as the pathway is an attractive target for drug development. While isoprenoids are essential for all organisms, humans employ a completely different isoprenoid-biosynthesis pathway: the mevalonate pathway. The MEP pathway has been chemically and genetically validated to be essential in a number of important pathogens, including *P. falciparum* (Cassera *et al.*, 2004; Odom & Van Voorhis, 2010; Zhang *et al.*, 2011), *Toxoplasma gondii* (Nair *et al.*, 2011), *Mycobacterium tuberculosis* (Brown & Parish, 2008) and *Escherichia coli* (Kuzuyama *et al.*, 1999). Furthermore, isoprenoid production is of interest for a wide range of commercially important natural products such as pharmaceuticals and microbial biofuels. Thus, structural insight into the regulation of the MEP pathway is important for its implications in drug development and for metabolic engineering to enhance isoprenoid production for commercial interests.

PfHAD1 serves as an excellent candidate for structural studies of substrate binding and catalysis within the HAD superfamily because its cellular function has been defined. We found that changes in PfHAD1 confer malaria-parasite resistance to the antimalarial fosmidomycin by sequencing *P. falciparum* parasites lines that are fosmidomycin-resistant (Guggisberg *et al.*, 2014). Fosmidomycin is a small-molecule inhibitor of the MEP pathway. Fosmidomycin-resistant parasite strains are highly enriched in genetic changes in the *PfHAD1* locus that result in loss of PfHAD1 activity. These genetic changes correlate with a metabolic effect of increased cellular levels of MEP pathway intermediates. Complementation of a deleterious *PfHAD1* allele restores sensitivity to fosmidomycin. While the metabolic effects and a biological phenotype of PfHAD1 have been described, the substrate specificity and mechanism of catalysis for PfHAD1, or HADs generally, have not been well defined.

The essential structural element of HADs is a Rossmannoid fold characterized by a three-layered α/β sandwich comprised of repeating β–α units (Burroughs *et al.*, 2006; Seifried *et al.*, 2013). The HAD Rossmannoid fold is distinguished from other Rossmann folds by two motifs called the 'squiggle' and 'flap' elements (Burroughs *et al.*, 2006; Allen & Dunaway-Mariano, 2009). The squiggle element occurs immediately

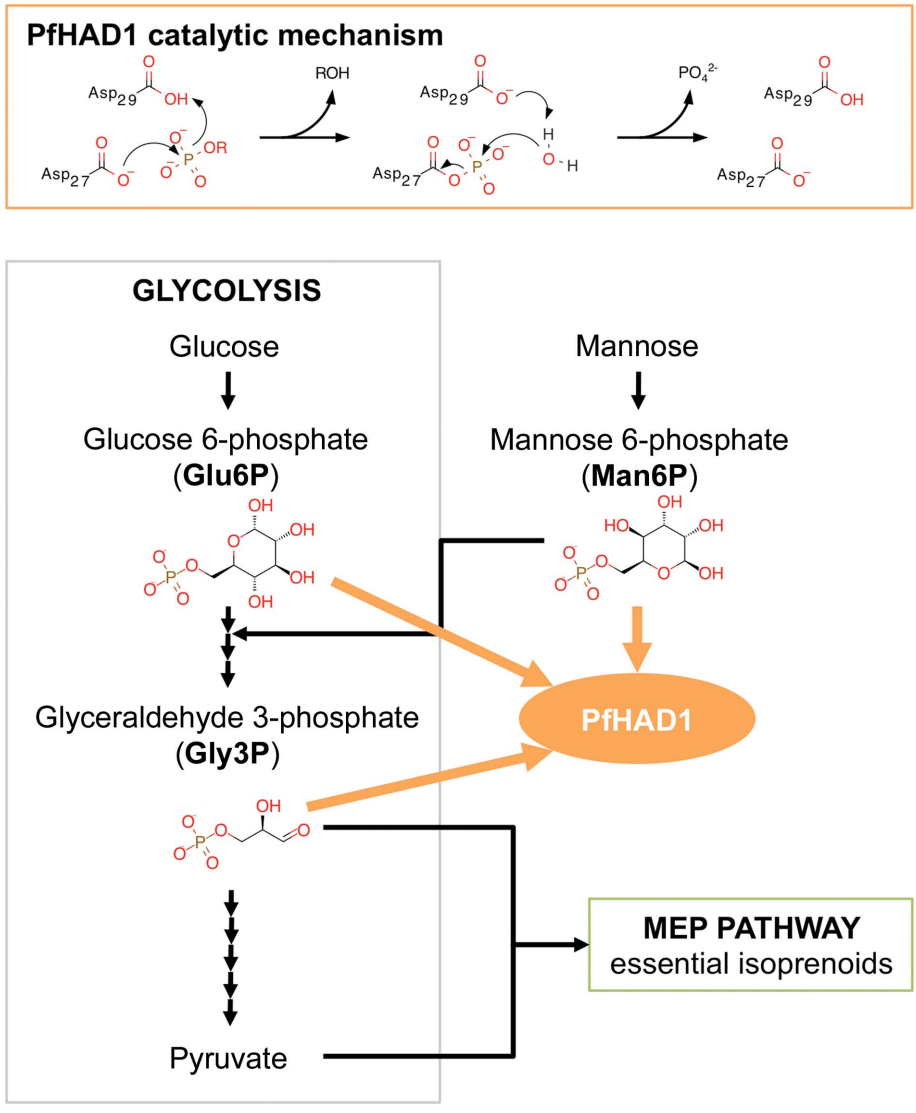


Figure 1
PfHAD1 regulates substrate availability to the MEP pathway and is capable of catalyzing the phosphorolysis of small, 3–6-carbon sugar monophosphates upstream of the MEP pathway, including mannose 6-phosphate (Man6P), glucose 6-phosphate (Glu6P) and glyceraldehyde 3-phosphate (Gly3P).

following the first β-strand of the Rossmannoid fold and assumes a nearly complete single-helical turn. The flap element is a β-hairpin downstream of the squiggle. While the core Rossmannoid folds of HAD enzymes have undergone minimal modifications, the introduction of a cap-insertion module has added a sophisticated means of substrate recognition and diversification of enzyme function by presenting new surfaces for interaction with substrates (Pandya *et al.*, 2014). HAD proteins can be classified into three general categories based on the topology and location of their cap insertions: C0, C1 or C2. The C0 element consists of short loops or β-strands that do not fold into its own structural unit, whereas the C1 element folds into an independently folded cap domain that is distinct from the catalytic core domain. The insertion in C0 and C1 HADs occurs in the middle of the β-hairpin of the flap motif. The C2 element also folds into its own distinct domain, but the cap insertion occurs in the linker following the third β-strand of the Rossmannoid fold.

In addition to providing substrate-specificity determinants, the cap module plays a role in solvent access and exclusion from the active site, which is a necessary aspect of catalysis *via* the two-step phosphoaspartyl-transferase mechanism (Seifried *et al.*, 2013; Peisach *et al.*, 2004). The C0 HADs, which lack a cap domain, primarily use large macromolecule substrates (Galburt *et al.*, 2002; Peisach *et al.*, 2004) or oligomerization (Parsons *et al.*, 2002; Biswas *et al.*, 2009; Lu *et al.*, 2009) as a means of substrate binding and solvent exclusion from the catalytic site. The C1 HADs employ extensive movement of the cap domain, which is dependent on winding and unwinding of the squiggle element, to form open and closed conformations (Allen & Dunaway-Mariano, 2009; Dai *et al.*, 2006, 2009). The role of cap movement in C2 HADs, such as PfHAD1, has not been extensively explored. It has been postulated that the squiggle–flap elements in C2 HADs also exhibit drastic movements to exclude solvent from the substrate-binding site (Burroughs *et al.*, 2006; Seifried *et al.*, 2013). However, to our knowledge evidence for this movement in C2 HADs has not been established.

Detailed structural analysis of the determinants for substrate ambiguity and specificity in HAD enzymes has not been investigated. To investigate the structural basis for substrate recognition and catalysis in *P. falciparum* PfHAD1, we constructed, purified and characterized a catalytically inactive PfHAD1-D27A mutant. We determined three substrate-bound crystal structures: those of PfHAD1-D27A complexed with mannose 6-phosphate (Man6P), glucose 6-phosphate (Glu6P) or glyceraldehyde 3-phosphate (Gly3P). Our crystallographic and enzymatic activity data provide insights into the structural determinants for substrate recognition and the role of C2 cap-domain movement in substrate binding, ordering and catalysis. We find that a large cap-domain movement results in sealing of the substrate-binding cavity. This movement is important for the enzymatic mechanism, as it positions cap residues in the correct orientation to bind and order substrates prior to catalysis. These results alter the paradigm that substrate specificity is imparted by the squiggle–flap motifs and is instead driven by cap closure. The structures of multiple substrate-bound complexes identified key specificity-determining residues within the cap. These specificity-determining residues are conserved in unique patterns within the HAD superfamily that aid in functional characterization. Understanding the PfHAD1 mechanism informs catalysis of the entire HAD superfamily, which regulates important metabolic pathways.

2. Materials and methods

2.1. Materials

All reagents were purchased from Sigma–Aldrich unless otherwise indicated.

2.2. Site-directed mutagenesis for PfHAD1 mutants

Wild-type PfHAD1 (WT-PfHAD1) was cloned as described previously (Guggisberg *et al.*, 2014). Site-directed mutagenesis to create the D27A allele was carried out using standard PCR-based techniques using the following primers: 5′-TGTTCCATCTAAAGCCGTAAAG-3′ and 5′-CTTTAC-GGCTTTAGATGGAACA-3′. Sites for ligation-independent cloning were added using the following primers: 5′-CTCACCACCACCACCACCATATGCACGAAATTGTA-GATAAGAA-3′ and 5′-ATCCTATCTTACTCACTTATA-TGTCACAGAATGTCTTCA-3′. The resulting PCR product was inserted into the BG1861 expression vector (Alexandrov *et al.*, 2004) to produce 6×His-PfHAD1. The construct sequence was verified by Sanger sequencing. Site-directed mutagenesis was performed to generate the V151A, V151L, E152A, L173A, E205A and E205W mutants.

2.3. Expression and purification

E. coli BL21(DE3) pLysS competent cells (Life Technologies) harboring the WT-PfHAD1 or mutant expression plasmids were grown at 37°C in LB medium containing final concentrations of 0.1 mg ml^{-1} ampicillin and 0.3 mg ml^{-1} chloramphenicol. Once the cells reached an OD_{600} of 0.6, they were induced with 1 m*M* IPTG for 3–4 h. After this period, the cells were harvested by centrifugation at 4000 rev min^{-1} for 10 min at 4°C. The cell pellets were suspended in 10 ml buffer *A* (50 m*M* Tris pH 8.0, 100 m*M* NaCl, 10 m*M* imidazole pH 8.0, 1 m*M* PMSF, 5 m*M* β-mercaptoethanol) per litre of LB medium and were stored at −80°C.

The cells were thawed in the presence of 0.25 mg ml^{-1} lysozyme and disrupted using sonication on ice for 60 s. The cell extract was obtained by centrifugation at 13 000 rev min^{-1} for 30 min at 4°C and was applied onto Nickel Rapid Run agarose beads (Goldbio) equilibrated with buffer *B* (50 m*M* Tris pH 8.0, 150 m*M* NaCl, 20 m*M* imidazole pH 8.0, 5 m*M* β-mercaptoethanol). Buffer *B* was used to wash the nickel column three times with five column volumes. After washing, the protein was eluted with five column volumes of buffer *C* (buffer *B* with 300 m*M* imidazole). The PHAD1 was further purified by gel chromatography using a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated with buffer *D* [10 m*M* Tris pH 8.0, 150 m*M* NaCl, 5 m*M*

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	PfHAD1-D27A + Man6P	PfHAD1-D27A + Glu6P	PfHAD1-D27A + Gly3P
Data collection			
Space group	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters			
a (Å)	77.60	77.70	77.60
b (Å)	44.50	44.50	44.60
c (Å)	84.50	84.80	84.10
$\alpha = \gamma$ (°)	90.00	90.00	90.00
β (°)	101.30	101.40	101.30
Resolution (Å)	20–1.80 (1.90–1.80)	20–1.90 (2.00–1.90)	20–2.00 (2.10–2.00)
R_{meas} (%)	6.5 (65.2)	8.4 (66.5)	11.1 (74.0)
$\langle I/\sigma(I)\rangle$	17.04 (2.45)	16.10 (2.42)	13.78 (2.20)
Completeness (%)	99.3 (99.6)	98.9 (98.3)	98.4 (97.7)
Multiplicity	3.68 (3.71)	3.77 (3.80)	3.79 (3.82)
Refinement			
Resolution (Å)	20–1.80	20–1.90	20–2.00
No. of reflections	52536	44805	38024
R_{work}/R_{free} (%)	18.83/21.91	17.89/21.95	17.32/22.10
No. of atoms			
Protein	4527	4512	4504
Mg	2	2	2
Ligand	21	21	15
Water	364	288	296
B factors (Å^2)			
Protein	29.90	29.80	34.30
Ligands	59.80	42.40	52.10
Water	30.20	28.80	32.50
Ramachandran			
Favored (%)	97.72	97.70	97.16
Allowed (%)	2.28	2.30	2.84
Outliers (%)	0.00	0.00	0.00
R.m.s. deviations			
Bond lengths (Å)	0.008	0.008	0.007
Bond angles (°)	1.198	1.187	0.969
PDB code	4zev	4zew	4zex

dithiothreitol (DTT)]. The fractions containing PfHAD1 were pooled and concentrated using a 10K molecular-weight cutoff Amicon centrifugal filter (Millipore).

2.4. Enzyme assays

Enzyme assays were performed as described previously (Guggisberg *et al.*, 2014). Briefly, general phosphatase activity was measured by monitoring the hydrolysis of *p*-nitrophenyl-phosphate (*p*NPP) to *p*-nitrophenyl. Reactions consisted of 10 m*M* *p*NPP (New England Biolabs), 50 m*M* Tris–HCl pH 7.5, 5 m*M* $MgCl_2$, 0.5 m*M* $MnCl_2$ and 2 µg enzyme. Phosphate cleavage of phosphorylated sugar substrates was monitored using the EnzChek Phosphatase Assay Kit (Life Technologies). Reactions contained 1 m*M* substrate and 400 ng enzyme. All substrates were purchased from Sigma–Aldrich. For all assays, reaction rates were calculated using *GraphPad Prism* software.

2.5. Crystallization

For crystallization, PfHAD1-D27A was concentrated to 20 mg ml^{-1} and incubated for 30 min on ice with 5 m*M* substrate for co-crystallization studies. Crystals of apo PfHAD1-D27A or PfHAD1-D27A grown in 5 m*M* substrate were obtained by vapor diffusion using hanging drops equilibrated at 18°C against 500 µl of a reservoir consisting of 0.1 *M* HEPES pH 7.5, 20–25% PEG 8000. PfHAD1-D27A crystals were cryoprotected with 0.1 *M* HEPES pH 7.5, 30% PEG 8000 and either no substrate for apo PfHAD1-D27A crystals or 5 m*M* substrate for co-crystals before flash-cooling under liquid nitrogen.

2.6. Data collection and processing

X-ray data were collected from a single crystal using a wavelength of 1 Å on beamline 4.2.2 of the Advanced Light Source, Berkeley, California, USA. Data were collected with the CMOS detector and were processed with *XDS* (Kabsch, 2010). All crystals had the same crystal packing and dimensions as the previously reported WT-PfHAD1 crystal (Guggisberg *et al.*, 2014). R and R_{free} flags were imported from the WT-PfHAD1 mtz file using *UNIQUEIFY* within the *CCP*4 package (Winn *et al.*, 2011).

2.7. Structure determination and refinement

Structure solution was performed within *PHENIX* (Adams *et al.*, 2010) by refinement with the previously solved WT-PfHAD1 crystal structure (PDB entry 4qjb). Subsequent iterated manual building/rebuilding and refinement of models were performed using *Coot* (Emsley & Cowtan, 2004) and *PHENIX* (Adams *et al.*, 2010), respectively. The *MolProbity* structure-validation server (Chen *et al.*, 2010) was used to monitor the refinement of the models. All final refined models have favorable crystallographic refinement statistics, as provided in Table 1. A stereo image of a representative region of the electron-density map is shown for each structure in Supplementary Fig. S1. Figures were generated and rendered in *PyMOL* (v.0.99rc6; Schrödinger).

2.8. Accession codes for the NCBI *BLAST* search

The PfHAD1 protein sequence was used as the query sequence to search for all nonredundant protein sequences in the NCBI database using the *BLASTP* 2.2.31+ (protein–protein *BLAST*) algorithm. *Plasmodium* (Taxid 5820) was excluded from the search. Sequence similarity was found to *Clostridium botulinum* WP_012342032, *Sebaldella termitidis* WP_012859625, *Atopococcus tabaci* WP_028273050, *Ilyobacter polytropus* WP_013387187, *Synergistes synergistes* sp. 3_1_syn1 WP_008709579, *Eucalyptus grandis* KCW79177, *Genlisea aurea* EPS68728, *Hammondia hammondi* XP_008889525, *Neospora caninum* Liverpool XP_003883270, *T. gondii* GT1 EPR64229, *E. coli* EOU46719, *M. tuberculosis* NP_218330, *C. reinhardtii* ADF43173 and *Arabidopsis thaliana* ABO38782. UniProt accession codes for *E. coli* HAD homologs are as follows: YbiV, P75792; YidA, P0A8Y5; YbhA, P21829; YbjI, P75809; YigL, P27848; OtsB, P31678; Cof, P46891.

3. Results

3.1. Generation of inactive PfHAD1-D27A for substrate-binding studies

Based on the sequence and structural analysis of WT-PfHAD1 (Guggisberg *et al.*, 2014), Asp27 was predicted to perform a nucleophilic attack on the phosphate group of PfHAD1 substrates. The PfHAD1-D27A mutant protein was generated and tested for catalytic activity towards a non-specific phosphatase substrate, *p*-nitrophenylphosphate (*p*NPP), and a representative panel of sugar phosphates, Man6P, Glu6P and Gly3P. These sugar phosphates represent six-carbon (6C; Man6P and Glu6P) and three-carbon (3C; Gly3P) sugar phosphates with different stereochemical properties that have biological relevance. Gly3P is one of the starting substrates for the MEP pathway, and Glu6P and Man6P are upstream precursors of Gly3P (Fig. 1). PfHAD1-D27A lacks phosphatase activity towards all compounds tested (Supplementary Fig. S2) and was employed for co-crystallization studies to obtain substrate-bound structures.

3.2. Overall structures of PfHAD1-D27A bound to substrates

Crystal structures of PfHAD1-D27A in complex with Man6P, Glu6P or Gly3P were determined. All structures are of high resolution (1.8–2.0 Å) sufficient to define small-molecule binding, and a complete summary of crystallographic refinement statistics is given in Table 1. The PfHAD1-D27A mutant has the same overall structure as WT-PfHAD1 (Fig. 2*a*), with the active site found at the interface between the core and cap domains. In each structure, two monomers were present in the asymmetric unit. Overlay of the two monomers by superposition of the core domains shows that PfHAD1 adopts a 'closed' and an 'open' conformation (Fig. 2*b*). Domain-movement analysis using *DynDom* (Taylor *et al.*, 2014; Hayward & Berendsen, 1998) reveals an 18° rotation between the two conformations about the two hinge loops (residues 104–105 and residues 210–214) connecting the core and cap domains.

For each of the substrate-bound structures, clear electron density corresponding to the substrate could be observed in the substrate-binding site of the closed conformation. The validity of the models was confirmed by inspecting the $F_o - F_c$ density prior to the addition of the substrate models (Fig. 3*a*) and inspecting the $2F_o - F_c$ electron-density maps after refinement with the substrate (Fig. 3*b*). In the open conformation, electron density corresponding to the phosphate group could be visualized, but the electron density for the orientation of the sugar moiety was unclear (Fig. 3*c*). Attempts to model in the substrate molecule were unsuccessful. Therefore, only the phosphate group was modeled in the open conformation (Fig. 3*d*).

3.3. PfHAD1 substrate-binding site

The active-site interactions in the co-crystal structures were examined more closely in order to investigate the structural basis for substrate binding and catalysis. HADs contain four highly conserved sequence motifs. These residues in PfHAD1 are as follows: motif I (Asp27 and Asp29), motif II (Thr61), motif III (Lys215) and motif IV (Asp238 and Asp242) (Figs. 4*a* and 4*b*). In the initial reaction, the Asp27 nucleophile (which is absent in our PfHAD1-D27A mutant) is positioned to attack the phosphoryl group (Figs. 4*a* and 4*b*). Asp29, which is positioned two residues away from Asp27, serves as a general acid/base residue to protonate the leaving sugar group. In the second reaction, Asp29 deprotonates a water molecule so that it can perform nucleophilic attack on the phosphoaspartyl intermediate, thus releasing free phosphate and restoring the enzyme to its native state (Fig. 1).

In addition to the conserved molecular interactions between the catalytic residues of the core domain and the phosphoryl group, residues in the cap domain make molecular

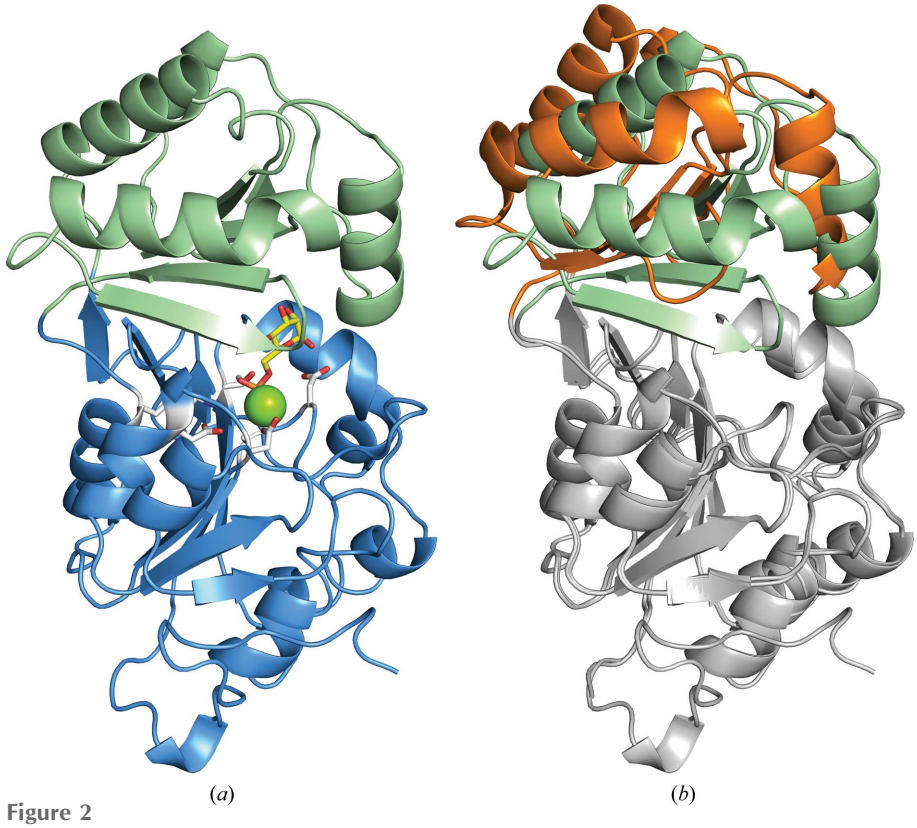


Figure 2
PfHAD1 binds and orders substrates in a closed conformation. (*a*) The overall PfHAD1-D27A + Man6P structure is shown here, with Man6P bound in the active site found at the interface between the core and cap domains. (*b*) Superimposition of the core domains of chain *A* and chain *B* reveals a shift in the cap domain relative to the core domain (gray), resulting in a closed (green) and an open (orange) conformation.

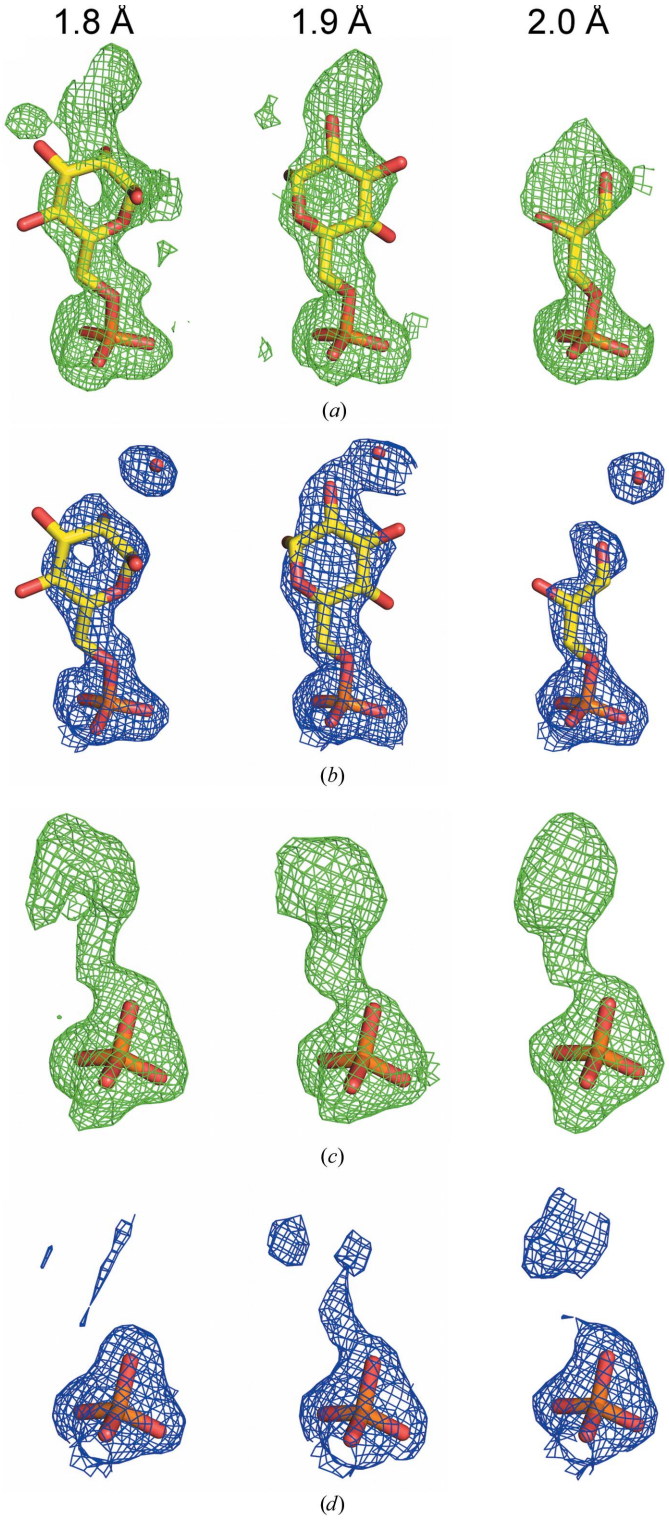


Figure 3
Substrates bind to the active site of PfHAD1-D27A. (*a*) The $F_o - F_c$ map (contoured at 2.5σ) of the substrate-binding site of the closed conformations prior to modeling of the substrate (from left to right: Man6P, Glu6P, Gly3P). Clear electron density can be observed for the entire substrate. (*b*) The $2F_o - F_c$ map (contoured at 1.0σ) of the closed substrate-binding site after refinement with the substrate model. (*c*) The $F_o - F_c$ map (contoured at 2.5σ) of the open substrate-binding site prior to modeling of the substrate. Electron density for the entire substrate is ambiguous. (*d*) The $2F_o - F_c$ map (contoured at 1.0σ) of the open substrate-binding site after refinement with a phosphate-group model.

interactions with the sugar moiety of the substrates in the closed conformation (Figs. 4*c*, 4*d* and 4*e*). In this conformation, the substrate-binding site is closed off and free exchange of water molecules with the bulk solvent is prevented, thus favoring the aspartate-based nucleophilic reaction. The open conformation, on the other hand, exposes the active site to bulk solvent, allowing substrate release. Owing to the shift of the cap domain away from the core domain in this open conformation, no molecular interactions are made between the sugar moiety of the substrates and the substrate-binding residues of the cap domain.

A comparison of the substrate-binding site in each of the three crystal structures reveals that Man6P, Glu6P and Gly3P all make contacts with cap residues Glu152, Thr201, Phe202 and Tyr205 (Figs. 4*c*, 4*d* and 4*e*). In addition, the larger 6C sugar phosphates (Man6P and Glu6P) make further contacts with cap residues Val151, Leu173, Glu207 and Arg63 from the catalytic domain (Figs. 4*c*, 4*d* and 4*e*). Owing to their different stereochemistry, slightly different interactions are made by Man6P and Glu6P. These contacts are consistent with the increased enzyme activities observed toward Man6P and Glu6P over Gly3P (Supplementary Fig. S2). Thus, distinct interactions for 6C and 3C sugar phosphates form the basis for substrate binding and specificity.

3.4. The cap domain contains a key substrate-recognition element

Although movement in the squiggle and flap elements has previously been postulated to be important for substrate binding and solvent exclusion in C2 HADs (Burroughs *et al.*, 2006), no movement in these elements was observed between the open and closed conformations of PfHAD1. In our structures, the flap element does not form a complete β-hairpin, but rather a simple β-turn. The *B*-factor values for the two monomers were examined in order to identify whether these regions demonstrated high flexibility and disorder in our structures. We observed average *B*-factor values for the squiggle and flap elements in all three structures, indicating that these elements are ordered in our structures (Figs. 5*a* and 5*a* and Supplementary Fig. S3). However, we observed abnormally high *B*-factor values in an α-helical region of the cap domain that is important for substrate binding. Glu152 of this region has the highest *B*-factor value of ∼110 Å (Fig. 4*a* and Supplementary Fig. S3*a*). In contrast, this region becomes ordered in the closed conformation, where the cap residues make direct interactions with the sugar moieties (Fig. 5*b* and Supplementary Fig. S3*b*). Glu152 makes a conserved hydrogen-bond interaction with an active-site water molecule in all of the substrate-bound structures (Figs. 4*c*, 4*d* and 4*e*). This water molecule hydrogen-bonds to sugar moieties of the substrates and helps to orient them in the active site. To test the relevance of Glu152 in catalysis, we generated a PfHAD1-E152A variant protein. Loss of Glu152 significantly lowers the enzyme activities equally for all substrates (Fig. 5*c*), demonstrating its important role in substrate recognition.

3.5. Substrate specificity informs the HAD superfamily

Val151, Glu152, Leu173, Thr201, Phe202, Tyr205 and Glu207 are cap-domain residues that bind substrates and enable specificity for PfHAD1. Having identified the structural determinants for substrate binding and specificity, we asked whether these determinants define the large superfamily of HAD enzymes. We examined the conservation of these substrate-binding residues in an unbiased sampling of the ten closest PfHAD1 homologs identified solely by sequence similarity.

Sequence alignment of these PfHAD1 homologs demonstrates that they are highly conserved in their catalytic core domains, but that their sequences diverge in their cap domains (Fig. 6*a* and Supplementary Fig. S4*a*). This is expected as our structural analysis indicates that the cap domains impart substrate specificity. Within these sequences, Glu207 is invariant and Leu173 is almost invariant, with substitutions to Met, Ile or Val in certain HADs (Fig. 6*b*). These two residues therefore play a conserved role in substrate binding across HADs, but a minor role in specificity. Strikingly, analysis of the

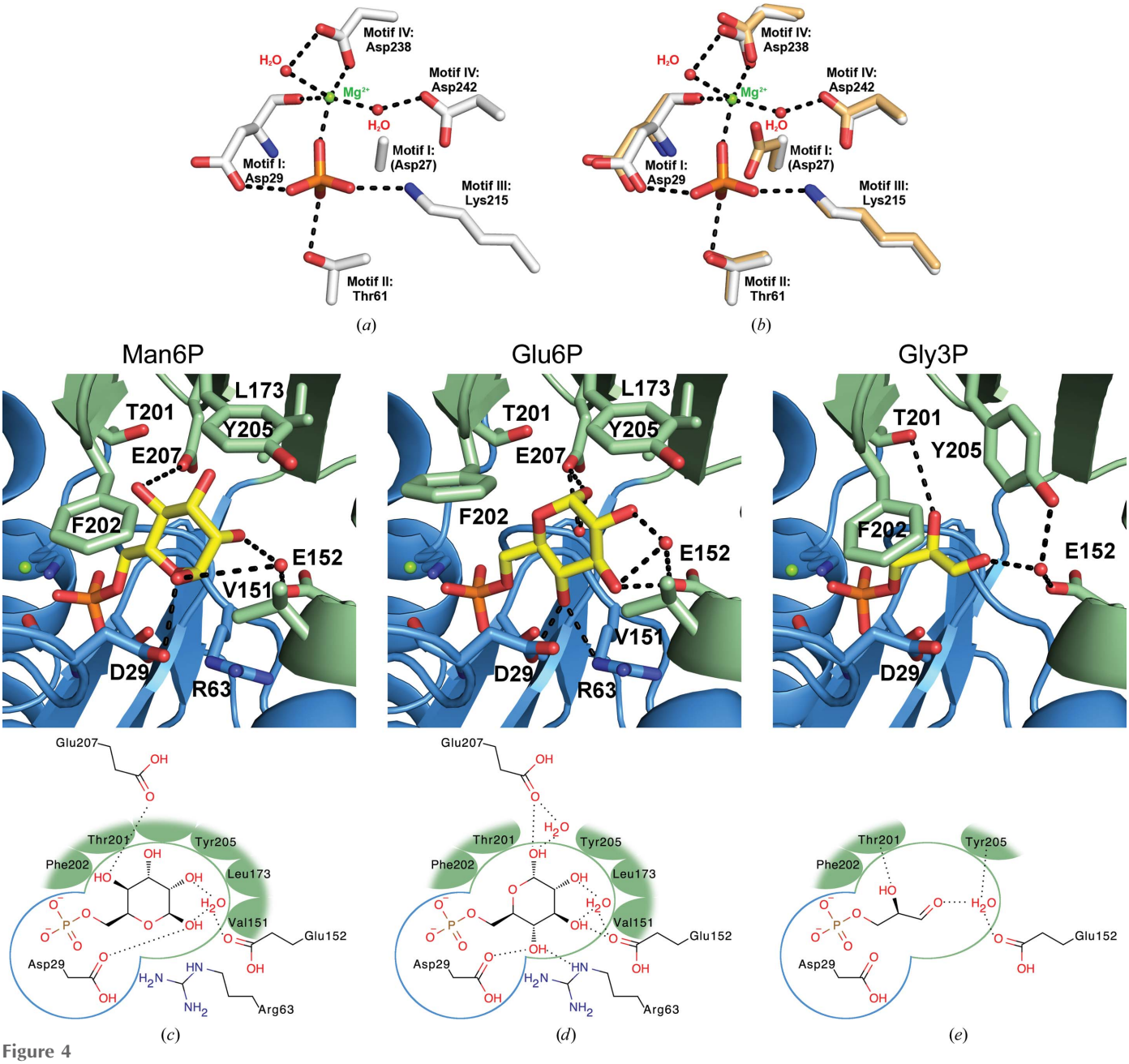


Figure 4

Distinct interactions for six-carbon and three-carbon sugar phosphates form the basis for substrate binding and specificity. (*a*) The conserved HAD motifs I–IV make interactions with a magnesium ion, a conserved water molecule and the phosphoryl group. (*b*) An overlay of active-site residues from the WT-PfHAD1 structure (colored light orange; PDB entry 4qjb) shows that the substrate-binding mode is conserved in the PfHAD1-D27A mutant. (*c*–*e*) Detailed van der Waals (colored green) and hydrogen-bond (dashed lines) interactions are shown between cap residues and the sugar moiety of each substrate: (*c*) Man6P, (*d*) Glu6P, (*e*) Gly3P. For simplicity, hydrogen-bond interactions between the conserved HAD motifs I–IV and the substrate are not shown here.

remaining residues suggests that HAD enzymes can be subdivided into two groups. In the first subdivision (Fig. 6*b*, blue boxes) the hydroxyl group of Thr201 and the aromatic side chain of Tyr205 are well conserved, while Val151, Glu152 and Phe202 are variable. The defining features of the second subdivision (Fig. 6*b*, red boxes) include an invariant Tyr at residue 151, Ala/Pro at residue 201, Val/Ile at residue 202 and Met/Asn at residue 205. Therefore, while the overall sequence conservation in the cap domains is lower that the catalytic core domains, patterns of sequence conservation and divergence are observed in cap residues that are important for substrate recognition. These patterns are only evident through the structural definition of substrate recognition presented in this study.

The analysis above used an unbiased sampling of proteins similar to PfHAD1. As PfHAD1 is a regulator of the MEP pathway in *P. falciparum* (Guggisberg *et al.*, 2014), we examined PfHAD1 homologs in other important pathogens and model organisms that also employ the MEP pathway: *E. coli*, *M. tuberculosis*, *A. thaliana* and *C. reinhardtii* (Supplementary Fig. S4*b*). All of these species employ the MEP pathway, as determined by the presence of DXR, the first committed enzyme of the pathway. Again, similar patterns of conservation and variance emerge within these selected PfHAD1 homologs (Fig. 6*c*). Val151 and Leu173 both provide hydrophobic contacts to the substrates and are important for sealing the substrate-binding cavity. However, Leu173 is strongly conserved while Val151 is variant (Figs. 5*b* and 5*c* and above). We mutated each of these residues to alanine to compare the effects of nonconserved and conserved residues on catalysis (Fig. 6*d*). Mutation of the conserved Leu173 resulted in an enzyme that was essentially inactive to all three substrates. In contrast, mutation of the nonconserved Val151 retained ∼75% of the wild-type activity for all three substrates. These results are consistent with the plasticity at Val151 and other variant residues to allow altered substrate specificity of HAD proteins.

None of the homologs from the analyses above have known substrate preferences except for *E. coli* YidA. The patterns of conserved and substituted binding residues identified by this structural work provide a framework for the future determination of substrate preferences for related HAD enzymes. To evaluate the validity of this framework, we examined the C2-type *E. coli* HAD enzymes YbiV, YidA, YbhA, YbjI, YigL, OtsB, Cof and NagD, as the substrate preferences of each of these enzyme is well defined (Kuznetsova *et al.*, 2006); C0 and C1 HAD enzymes were omitted from this analysis as they have different domain architectures. YbiV and YidA both catalyze dephosphorylation of 3–6-carbon sugar phosphates similar to PfHAD1, with YbiV having greatest activity against fructose 1-phosphate and YidA preferring erythrose 4-phosphate. As predicted, we find that YbiV and YidA have similar

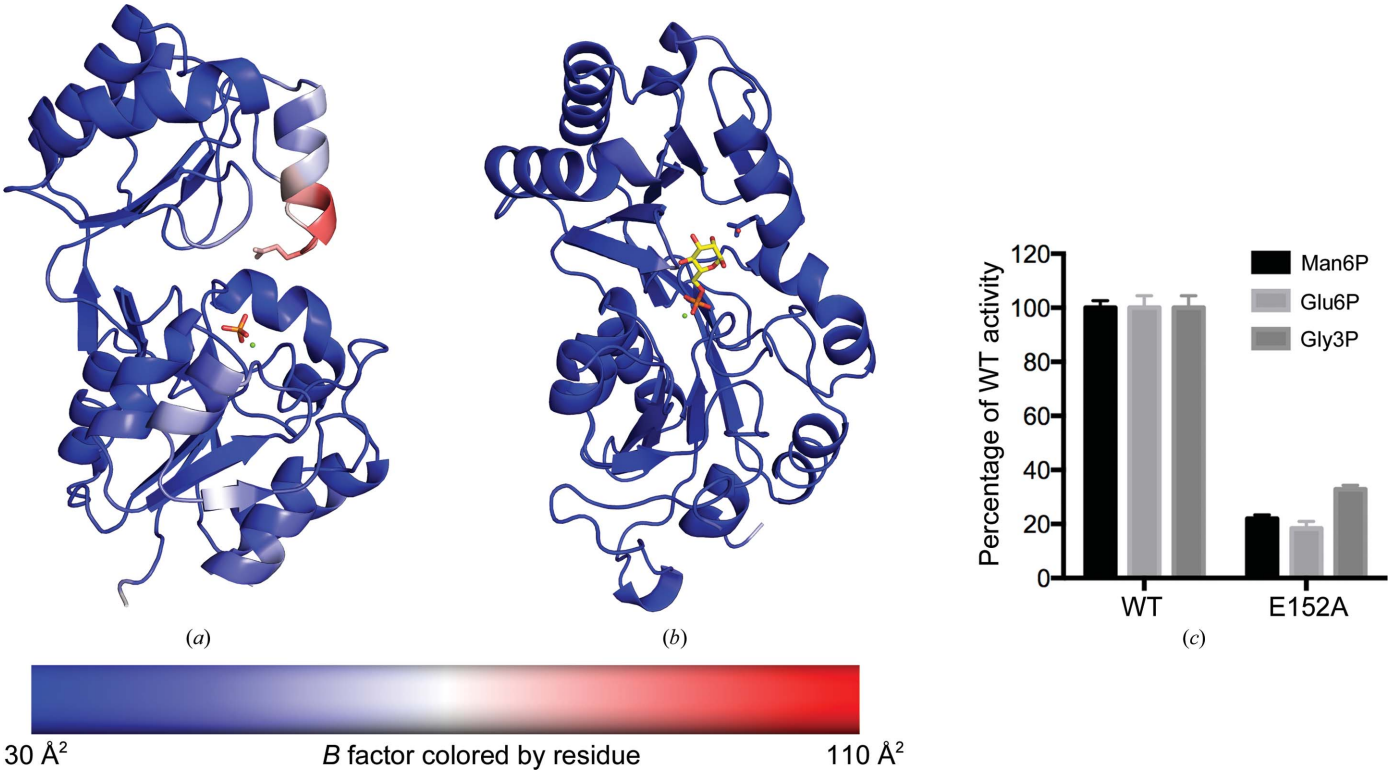


Figure 5
The cap domain contains a flexible substrate-recognition element that is ordered upon substrate binding and cap closure. (*a*) In the open conformation for Man6P, the region of the cap domain that is important for substrate recognition has high *B*-factor values, indicating disorder and flexibility in this region. (*b*) In the closed conformation for Man6P, this region becomes ordered, as reflected by the low *B*-factor values. (*c*) Enzyme activities for WT-PfHAD1 and PfHAD1-E152A are shown, normalized as a percentage of the WT-PfHAD1 activity for each substrate. Displayed are the means ± standard error of the mean of enzyme activity from at least three independent experiments.

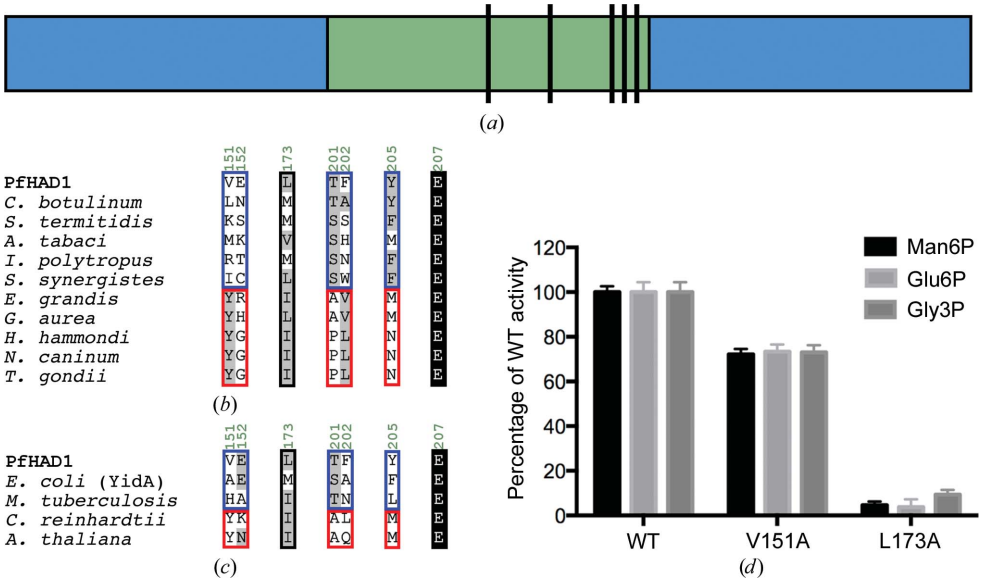


Figure 6
Sequence alignment of substrate-binding cap-domain residues in PfHAD1 and PfHAD1 homologs from other organisms. (*a*) PfHAD1 homologs in other organisms are highly conserved in their catalytic core domains (blue) but diverge in their cap-domain sequences (green). The black lines denote the five segments comprising the seven residues of the cap domain in PfHAD1 that are important for substrate binding. (*b*) The substrate-binding cap residues in the ten closest PfHAD1 homologs are shown. The blue and red boxes denote two subdivisions of PfHAD1 homologs. (*c*) The substrate-binding cap residues in PfHAD1 homologs from model organisms are shown. (*d*) Enzyme activities for WT-PfHAD1, PfHAD1-V151A and PfHAD1-L173A are shown, normalized as a percentage of the WT-PfHAD1 activity for each substrate. Displayed are the means ± standard error of the mean of enzyme activity from at least three independent experiments.

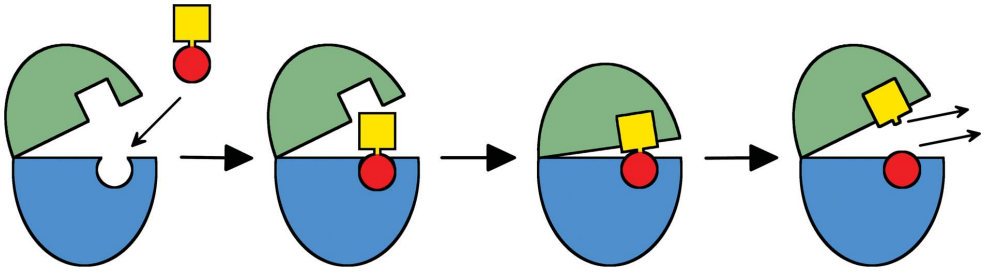


Figure 7
Model for catalysis in PfHAD1. The open conformation of PfHAD1 allows the substrate to access the active site and bind the conserved catalytic residues with the phosphoryl group. Subsequent cap closure prevents the free exchange of water molecules with bulk solvent and allows ordering of the substrate for catalysis. Opening of the cap allows solvent access to restore PfHAD1 to its native state.

patterns of conservation/divergence as PfHAD1: variable sequences at positions 151/152, Leu/Met at position 173, Tyr/Ser at position 201, Tyr/Phe at position 205 and Asp/Glu at position 207 (Supplementary Fig. S4*c*). In contrast, YbhA, YbjI, YigL, OtsB, Cof and NagD utilize structurally different phosphorylated metabolites. YbhA, YigL and Cof have greatest activity against pyridoxal 5-phosphate, YbjI is most active against flavin mononucleotides, OtsB has a preference for trehalose 6-phosphate and NagD is most active against adenosine triphosphate. These *E. coli* homologs diverge dramatically from PfHAD1 at the substrate-binding residues (Supplementary Fig. S4*c*), further validating the correlation between structure and substrate preference.

4. Discussion

In this study, we have solved co-crystal structures of the sugar phosphatase PfHAD1 bound to diverse sugar-phosphate substrates in order to investigate the structural determinants for substrate recognition and specificity. By investigating molecular interactions within PfHAD1 bound to multiple different substrates, our study is the first structural study to examine how the HAD enzyme family is able to remain promiscuous with substrate utilization: a common feature of the HAD superfamily members, which is otherwise unusual among enzymes that recognize metabolites. Our results demonstrate that PfHAD1 utilizes conserved residues in the catalytic domain to bind the phosphoryl group of sugar-phosphate substrates. Substrate specificity is determined by residues in the cap domain that make interactions with the sugar moiety, thereby defining which substrates PfHAD1 can utilize.

Prior to this study, the role of cap movement in C2 HADs was unclear. It was speculated that C2 HADs undergo structural changes in the squiggle and flap elements rather than the cap domain (Allen & Dunaway-Mariano, 2009; Burroughs *et al.*, 2006) to allow catalysis to occur. However, no direct biochemical or structural evidence for this squiggle–flap movement has been reported. The only other structural clue was reported for a cyanobacterial sucrose-phosphatase, for which an open and closed conformation of the cap domain in two different crystal forms was observed (Fieulaine *et al.*, 2005).

Here, we have observed open and closed conformations within the same crystal structures, and we find that these changes coordinate substrate binding and ordering. The open and closed conformations are likely to represent physiologically relevant states of PfHAD1 that are essential for substrate binding and catalysis (Fig. 7). The open conformation of the enzyme allows a substrate molecule to bind the catalytic residues of the core domain with its phosphoryl group. In this open state, we see clear electron density for the phosphoryl

group but not the sugar moiety that is facing bulk solvent. Upon inspection of the *B*-factor values in these structures, we observe that a substrate-binding element in the cap domain is highly flexible and disordered. We have also captured the subsequent closed conformation of PfHAD1 in our structures. In this closed state, a substrate-binding cavity is formed at the interface between the core and cap domains that prevents the free exchange of water molecules with bulk solvent. The cap domain orders the substrate through polar and nonpolar interactions with the sugar moiety of the substrate, and low *B* factors are observed for the entire protein. This closed state provides an environment that favors an aspartate-based nucleophilic attack to occur. Subsequent opening of the cap would facilitate product release and allow the enzyme to be restored to its native state.

We identified distinct residues that mediate substrate binding and specificity for PfHAD1. Analysis of these residues in HAD homologs identified patterns of conservation and variance that subdivide the superfamily and may predict their substrate specificities. This analysis suggests that the subdivisions of HAD homologs may represent similar substrate-utilization profiles and/or biological functions *in vivo*. While mutation of the conserved cap residues are detrimental to enzyme function in these HADs, the variable cap residues are more suitable for changes to alter substrate specificity or for metabolic engineering. Understanding which residues to manipulate enables rational protein-engineering approaches to alter the substrate-utilization profiles of these enzymes. Future biochemical and structural studies of these PfHAD1 homologs will further our understanding of the substrate specificity and biological functions of these additional HAD enzymes.

HAD superfamily members are widespread in biology, but few subfamilies have experimentally defined biological roles. We previously found that PfHAD1 regulates substrate availability to the MEP pathway for isoprenoid precursor biosynthesis and that loss of PfHAD1 function has dramatic effects on the levels of MEP pathway intermediates (Guggisberg *et al.*, 2014). PfHAD1 homologs in other organisms are restricted to phyla that also employ the MEP pathway, suggesting that PfHAD1-like HADs are also regulators of the MEP pathway in these other organisms. Thus, our studies are an important first step in the development of chemical tools to disrupt or alter HAD enzyme functions. These tools will allow us to probe the biological and metabolic functions of HAD enzymes in living cells. As the MEP pathway is essential in parasites and pathogenic bacteria, defining the regulation of this pathway by HAD enzymes may lead to novel avenues for therapeutic intervention.

Acknowledgements

We thank J. Nix and ALS beamline 4.2.2 (contract DE-AC02-05CH11231) for assistance with X-ray data collection. This work was supported by the Children's Discovery Institute of Washington University and St Louis Children's Hospital (MD-LI-2011-171 to AO and NT), NIH/NIAID R01AI103280 (to AO), a March of Dimes Basil O'Connor Starter Scholar Research Award (to AO) and a Doris Duke Charitable Foundation Clinical Scientist Development award (to AO). JP is supported by a Sondra Schlesinger Graduate Student Fellowship and a Stephen I. Morse Graduate Student Fellowship from Washington University. AMG is supported by an NIGMS Training grant (5T32GM007067) and a Monsanto Excellence Fund Graduate Fellowship. JP and AMG contributed equally to this work and are co-first authors. ARO and NHT also contributed equally to this work and are co-senior authors. All authors were involved in experiment design, data interpretation and writing of the manuscript.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Alexandrov, A., Vignali, M., LaCount, D. J., Quartley, E., de Vries, C., De Rosa, D., Babulski, J., Mitchell, S. F., Schoenfeld, L. W., Fields, S., Hol, W. G. J., Dumont, M. E., Phizicky, E. M. & Grayhack, E. J. (2004). *Mol. Cell. Proteomics*, **3**, 934–938.
Allen, K. N. & Dunaway-Mariano, D. (2004). *Trends Biochem. Sci.* **29**, 495–503.
Allen, K. N. & Dunaway-Mariano, D. (2009). *Curr. Opin. Struct. Biol.* **19**, 658–665.
Biswas, T., Yi, L., Aggarwal, P., Wu, J., Rubin, J. R., Stuckey, J. A., Woodard, R. W. & Tsodikov, O. V. (2009). *J. Biol. Chem.* **284**, 30594–30603.
Brown, A. C. & Parish, T. (2008). *BMC Microbiol.* **8**, 78.
Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D. & Aravind, L. (2006). *J. Mol. Biol.* **361**, 1003–1034.
Cassera, M. B., Gozzo, F. C., D'Alexandri, F. L., Merino, E. F., del Portillo, H. A., Peres, V. J., Almeida, I. C., Eberlin, M. N., Wunderlich, G., Wiesner, J., Jomaa, H., Kimura, E. A. & Katzin, A. M. (2004). *J. Biol. Chem.* **279**, 51749–51759.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Collet, J.-F., Gerin, I., Rider, M. H., Veiga-da-Cunha, M. & Van Schaftingen, E. (1997). *FEBS Lett.* **408**, 281–284.
Collet, J.-F., Stroobant, V. & Van Schaftingen, E. (2002). *Methods Enzymol.* **354**, 177–188.
Dai, J., Finci, L., Zhang, C., Lahiri, S., Zhang, G., Peisach, E., Allen, K. N. & Dunaway-Mariano, D. (2009). *Biochemistry*, **48**, 1984–1995.
Dai, J., Wang, L., Allen, K. N., Radstrom, P. & Dunaway-Mariano, D. (2006). *Biochemistry*, **45**, 7818–7824.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fieulaine, S., Lunn, J. E., Borel, F. & Ferrer, J.-L. (2005). *Plant Cell*, **17**, 2049–2058.
Fortpied, J., Maliekal, P., Vertommen, D. & Van Schaftingen, E. (2006). *J. Biol. Chem.* **281**, 18378–18385.
Galburt, E. A., Pelletier, J., Wilson, G. & Stoddard, B. L. (2002). *Structure*, **10**, 1249–1260.
Guggisberg, A. M., Park, J., Edwards, R. L., Kelly, M. L., Hodge, D. M., Tolia, N. H. & Odom, A. R. (2014). *Nature Commun.* **5**, 4467.
Hayward, S. & Berendsen, H. J. (1998). *Proteins*, **30**, 144–154.
Hunter, S. *et al.* (2012). *Nucleic Acids Res.* **40**, D306–D312.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kim, Y., Yakunin, A. F., Kuznetsova, E., Xu, X., Pennycooke, M., Gu, J., Cheung, F., Proudfoot, M., Arrowsmith, C. H., Joachimiak, A., Edwards, A. M. & Christendat, D. (2004). *J. Biol. Chem.* **279**, 517–526.
Koonin, E. V. & Tatusov, R. L. (1994). *J. Mol. Biol.* **244**, 125–132.
Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V., Borozan, I., Carmel, L., Wolf, Y. I., Mori, H., Savchenko, A. V., Arrowsmith, C. H., Koonin, E. V., Edwards, A. M. & Yakunin, A. F. (2006). *J. Biol. Chem.* **281**, 36149–36161.

Kuzuyama, T., Takahashi, S. & Seto, H. (1999). *Biosci. Biotechnol. Biochem.* **63**, 776–778.
Lu, Z., Wang, L., Dunaway-Mariano, D. & Allen, K. N. (2009). *J. Biol. Chem.* **284**, 1224–1233.
Nair, S. C., Brooks, C. F., Goodman, C. D., Sturm, A., Strurm, A., McFadden, G. I., Sundriyal, S., Anglin, J. L., Song, Y., Moreno, S. N. J. & Striepen, B. (2011). *J. Exp. Med.* **208**, 1547–1559.
Odom, A. R. & Van Voorhis, W. C. (2010). *Mol. Biochem. Parasitol.* **170**, 108–111.
Pandya, C., Farelli, J. D., Dunaway-Mariano, D. & Allen, K. N. (2014). *J. Biol. Chem.* **289**, 30229–30236.
Parsons, J. F., Lim, K., Tempczyk, A., Krajewski, W., Eisenstein, E. & Herzberg, O. (2002). *Proteins*, **46**, 393–404.
Passariello, C., Forleo, C., Micheli, V., Schippa, S., Leone, R., Mangani, S., Thaller, M. C. & Rossolini, G. M. (2006). *Biochim. Biophys. Acta*, **1764**, 13–19.
Peisach, E., Selengut, J. D., Dunaway-Mariano, D. & Allen, K. N. (2004). *Biochemistry*, **43**, 12770–12779.
Rangarajan, E. S., Proteau, A., Wagner, J., Hung, M.-N., Matte, A. & Cygler, M. (2006). *J. Biol. Chem.* **281**, 37930–37941.
Rinaldo-Matthis, A., Rampazzo, C., Reichard, P., Bianchi, V. & Nordlund, P. (2002). *Nature Struct. Biol.* **9**, 779–787.
Seifried, A., Schultz, J. & Gohla, A. (2013). *FEBS J.* **280**, 549–571.
Taylor, D., Cawley, G. & Hayward, S. (2014). *Bioinformatics*, **30**, 3189–3196.
Titz, B., Häuser, R., Engelbrecher, A. & Uetz, P. (2007). *FEMS Microbiol. Lett.* **270**, 49–57.
Wang, L., Lu, Z., Allen, K. N., Mariano, P. S. & Dunaway-Mariano, D. (2008). *Chem. Biol.* **15**, 893–897.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wu, J. & Woodard, R. W. (2003). *J. Biol. Chem.* **278**, 18117–18123.
Zhang, B., Watts, K. M., Hodge, D., Kemp, L. M., Hunstad, D. A., Hicks, L. M. & Odom, A. R. (2011). *Biochemistry*, **50**, 3570–3577.