Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a haloacid dehalogenase superfamily phosphatase PH1421 from *Pyrococcus horikoshii* OT3: oligomeric state and thermoadaptation mechanism

Hitoshi Yamamoto, Koji Takio, Mitsuaki Sugahara and Naoki Kunishima*

RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan

Correspondence e-mail: kunisima@spring8.or.jp

Received 1 April 2008
Accepted 12 August 2008

PDB Reference: PH1421, 1wr8, r1wr8sf.

PH1421 from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3 is a hypothetical protein belonging to the haloacid dehalogenase (HAD) superfamily. To gain insight into its biological function and thermostabilization mechanism, the crystal structure of PH1421 has been determined at 1.6 Å resolution. The crystallographic asymmetric unit contains a homodimer. The monomeric protomer is composed of two distinct domains, a small cap domain and a large core domain, which agrees well with the typical domain organization of HAD subfamily II. Based on structure-based amino-acid sequence alignment and enzymatic analysis, PH1421 is suggested to be a magnesium-dependent phosphatase that is similar to the dimeric HAD phosphatase TA0175 from the mesothermophilic archaeon *Thermoplasma acidophilum*. Further comparison between the crystal structures of PH1421 and TA0175 revealed a marked structural similarity in the interprotomer dimer association. The common dimer interface with interprotomer twofold symmetry is characterized by a well conserved hydrophobic core consisting of the $\beta1$–$\alpha1$ loop and helices $\alpha1$ and $\alpha2$ of the core domain and additional contacts including the $\beta7$–$\beta8$ loop of the cap domain, which constitutes part of the putative active site of the enzyme. Several factors that potentially contribute to the higher thermal stability of PH1421 were identified: (i) an increase in intraprotomer hydrophobic interactions, (ii) a decrease in denaturation entropy from amino-acid composition and (iii) an increased number of intraprotomer ion pairs. These results suggest that the PH1421 protomer itself has an intrinsically higher thermal stability when compared with the mesothermophilic orthologue TA0175.

1. Introduction

Phosphoryl-transfer reactions are widespread in many aspects of biological processes in all organisms. Enzymes in the haloacid dehalogenase (HAD) superfamily, also known as the L-2-haloacid dehalogenase family, catalyze diverse phosphoryl-transfer reactions in a magnesium-dependent manner using aspartate as a nucleophile (Aravind *et al.*, 1998; Collet *et al.*, 1998). More than 40 crystal structures of HAD proteins have been reported (for a review, see Burroughs *et al.*, 2006). Of these proteins, functionally characterized HAD enzymes have so far shown a clear correlation between their structures and enzymatic functions; the structurally conserved large core domain and the small variable cap domain are required for phosphatase catalysis and for substrate recognition, respectively.

The core domain contains the conserved residues for phosphatase activity (Allen & Dunaway-Mariano, 2004). This

domain has a well conserved α/β structure which is similar to the Rossmann fold. It contains three conserved motifs: I, II and III (Fig. 1). Motif I is described as D*X*D*X*(T/V) (where *X* denotes any residue). The first invariant aspartate acts as a nucleophile in the reaction and becomes phosphorylated before transferring the phosphoryl group (Collet *et al.*, 1998). The second aspartate acts as an acid/base catalyst (Lahiri *et al.*, 2002). Motif II is described as (S/T)G. These residues make hydrogen bonds to the phosphoryl O atoms of the substrate and help to orient it for nucleophilic attack (Wang *et al.*, 2001). Motif III is described as K-$(X)_n$-(G/S)(D/S)*XXX*(D/N). Of these residues, the conserved lysine neutralizes the negative charge of the phosphoryl groups during the reaction and the other residues are involved in the coordination of divalent metal ions (Wang *et al.*, 2001).

The cap domain, on the other hand, is structurally variable. It is inserted into the core domain at a typical position. According to the position of insertion, the HAD superfamily is divided into three subfamilies: I, II and III (Selengut, 2001). In subfamily I, the cap domain is inserted between motifs I and II in the core domain and is α-helical in structure; this subfamily includes the phosphoserine phosphatase from *Methanococcus jannaschii* (Wang *et al.*, 2001; PDB code 1f5s), the β-phosphoglucomutase from *Lactococcus lactis* (Lahiri *et al.*, 2003; PDB code 1o08) and the phosphonoacetaldehyde hydrolase from *Bacillus cereus* (Morais *et al.*, 2000; PDB code 1fez). In subfamily II, the cap domain is located between motifs II and III in the core domain and has a mixed α/β-fold. This subfamily includes the sugar phosphatase BT4131 from *Bacteroides thetaiotaomicron* VPI-5482 (Lu *et al.*, 2005; PDB code 1ymq), the TM0651 phosphatase from *Thermotoga maritima* (Shin *et al.*, 2003; PDB code 1nf2), the Ybiv protein from *Escherichia coli* (Roberts *et al.*, 2005; PDB code 1rlo) and the 2-phosphoglycolate phosphatase from *Thermoplasma acidophilum* (Kim *et al.*, 2004; PDB code 1l6r). The subfamily III enzymes comprise the core domain only and lack the cap domain. This subfamily includes the phosphotyrosine phosphatase MDP-1 from mouse (Peisach *et al.*, 2004; PDB code 1u7o) and the T4 phage polynucleotide kinase (Galburt *et al.*, 2002; PDB code 1ltq). The cap domain in subfamilies I and II is located in the vicinity of the catalytic site and protects it from the bulk solution. It has been reported that the cap domain shows 'open–closed' interdomain movement during the catalytic cycle [see Burroughs *et al.* (2006) for subfamily I, and Lu *et al.* (2005) and Burroughs *et al.* (2006) for subfamily II]. Furthermore, the cap domain modulates the substrate

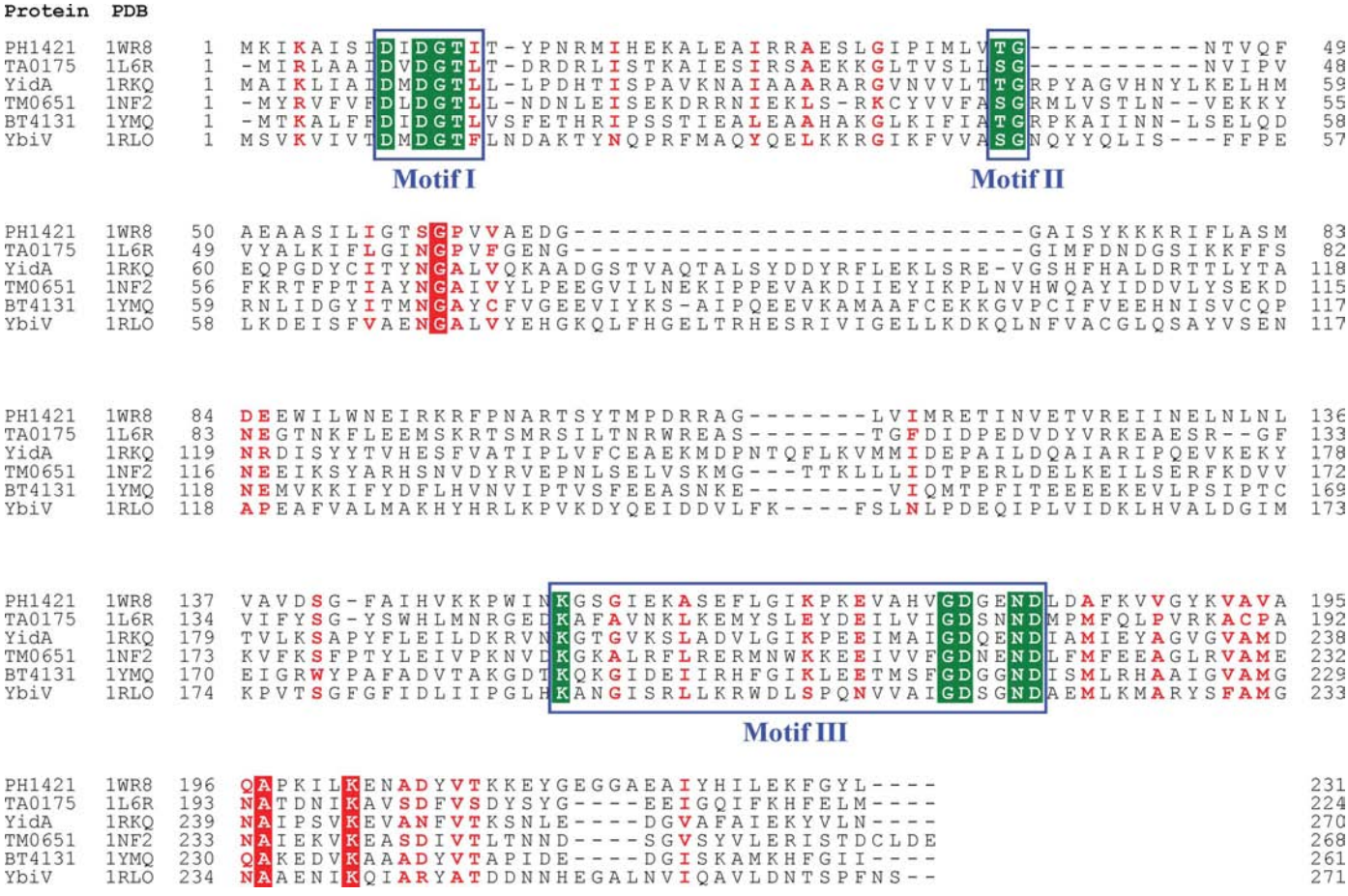


Figure 1
Sequence alignment of PH1421 and five HAD-family members belonging to subfamily II. The three signature motifs are boxed. Conserved active-site residues for phosphatase activity are highlighted in green. Other conserved residues are highlighted in red. This figure was produced using *ClustalW* (Chenna *et al.*, 2003).

recognition of these enzymes through the substrate-specificity loop (Tremblay *et al.*, 2006; Lahiri *et al.*, 2004; Lu *et al.*, 2005).

The source organisms of HAD proteins range from mesophiles to thermophiles. However, to date their thermoadaptation mechanism has not been fully discussed from a structural viewpoint. Therefore, comparing HAD-protein structures from organisms with different living temperatures may provide an insight into the thermoadaptation mechanism of these proteins. Here, we report the crystal structure of PH1421 from the hyperthermophilic archeon *Pyrococcus horikoshii* OT3 at 1.6 Å resolution. From structural and functional characterization, we successfully identified PH1421 as a dimeric magnesium-dependent phosphatase belonging to HAD subfamily II. Based on the three-dimensional information, the biological significance of the oligomeric state and the thermoadaptation mechanism will be discussed.

2. Materials and methods

2.1. Sample preparation

The 25.6 kDa monomeric protomer of PH1421 consists of 231 amino-acid residues. The plasmid encoding PH1421 (residues 1–231) was digested with *Nde*I and *Bgl*II and the fragment was inserted into the expression vector pET-11a (Novagen Inc.) linearized with *Nde*I and *Bam*HI. 2.3 l Luria–Bertani medium supplemented with ampicillin at a concentration of 100 µg ml^{-1} was inoculated with a single colony of *E. coli* BL21-Codon Plus (DE3)-RIL (Novagen Inc.) carrying the recombinant plasmid and grown at 310 K for 20 h with vigorous shaking. Cells were harvested and resuspended in 22 ml buffer solution containing 20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 5 m*M* 2-mercaptoethanol and 1 m*M* PMSF and sonicated. The cell lysate was centrifuged at 21 600*g* for 30 min at 277 K and the supernatant was then incubated at 363 K for 10 min. The heat-treated lysate was centrifuged at 21 600*g* for 30 min at 277 K. The supernatant was desalted with a HiPrep 26/10 column (GE Healthcare Biosciences Inc.) equilibrated with 20 m*M* Tris–HCl pH 8.0 (buffer *A*) and applied onto a Super Q Toyopearl 650M (Tosoh Inc.) anion-exchange column (80 ml) equilibrated with buffer *A* and eluted with a 240 ml linear gradient of 0–0.3 *M* NaCl. The fractions were analyzed by SDS–PAGE and those containing a 25.6 kDa protein were collected. The protein solution was desalted with a HiPrep 26/10 column equilibrated with buffer *A* and applied onto a Resource Q (GE Healthcare Biosciences Inc.) anion-exchange column equilibrated with buffer *A*. The protein was eluted with a 60 ml linear gradient of 0–0.3 *M* NaCl. The fractions were analyzed and collected as described above. The protein solution was desalted with a HiPrep 26/10 column equilibrated with 10 m*M* potassium phosphate pH 7.0 (buffer *B*) and applied onto a CHT 20-I (Bio-Rad Inc.) hydroxyapatite column equilibrated with buffer *B*. The protein was eluted with a 300 ml linear gradient of 0.01–0.5 *M* potassium phosphate pH 7.0. The fractions were analyzed and collected as described above. The protein solution was concentrated by ultrafiltration (Vivaspin, 10 kDa cutoff; Vivascience) and applied onto a HiLoad 16/60 Superdex 200 prep-grade column equilibrated with buffer *A* containing 200 m*M* NaCl. The fractions were analyzed and collected as described above. The purified protein solution was concentrated to 1.75 ml by ultrafiltration. The protein concentration was determined to be 37.1 mg ml^{-1} by measuring the UV absorption at 280 nm, considering a molar absorption coefficient of 29 085 M^{-1} cm^{-1}.

2.2. Dynamic light scattering

The oligomerization state of PH1421 was analyzed by dynamic light scattering using a DynaPro MS/X instrument (Protein Solutions Inc.). PH1421 at a concentration of 1.0 mg ml^{-1} in 20 m*M* Tris–HCl pH 8.0 and 200 m*M* NaCl was centrifuged at 21 600*g* for 30 min and the supernatant was used for data collection. Ten sets of scattering data were taken at 289 K for each sample. The regularization histogram was analyzed using the *DYNAMICS* software v.5.26.60 (Protein Solutions Inc.). The apparent molecular weight was calculated as 49.7 kDa, which is consistent with a dimeric state of PH1421 in solution.

2.3. Crystallization and heavy-atom derivatization

Crystals of PH1421 protein were grown by the oil-microbatch method (Chayen *et al.*, 1990) using a Nunc HLA plate (Nalge Nunc Instrumental). Initial screening of crystallization conditions was performed using the TERA automatic crystallization system (Sugahara & Miyano, 2002). In the optimized final condition, the crystallization drop was prepared by mixing 0.5 µl protein solution comprising 10.0 mg ml^{-1} PH1421, 200 m*M* NaCl and 20 m*M* Tris–HCl pH 8.0 and 0.5 µl precipitant solution comprising 25.5%(*w*/*v*) polyethylene glycol 4000, 0.17 *M* ammonium acetate, 15%(*w*/*v*) glycerol and 85 m*M* sodium acetate buffer pH 4.6. The crystallization drop was overlaid with a 1:1 mixture of silicone and paraffin oils, allowing slow evaporation of water in the drop, and stored at 291 K. Crystals grew in less than a week. A dysprosium-derivative crystal was prepared by soaking for 1 d at 291 K in precipitant solution containing 10 m*M* dysprosium(III) chloride hexahydrate.

2.4. Data collection

For both the native and dysprosium-derivative cases, crystals were flash-cooled in a nitrogen-gas stream at 100 K using the precipitant solution as a cryoprotectant. A three-wavelength MAD experiment was performed using a single dysprosium-derivative crystal on beamline BL26B1 of SPring-8, Japan (Ueno *et al.*, 2006). The dysprosium-derivative crystal diffracted X-rays to 2.0 Å resolution. A native data set was collected at 1.6 Å resolution. The statistics of data collection are summarized in Table 1. The *HKL*-2000 software package (Otwinowski & Minor, 1997) was used for data reduction and scaling. The native crystal belonged to space group $P2_12_12_1$, with unit-cell parameters a = 63.8, b = 81.8, c = 87.0 Å. Assuming two chains of PH1421 in the asymmetric

unit, the V_M value and solvent content were calculated as 2.2 Å^3 Da^{-1} and 44.5%, respectively.

2.5. Structure determination

Phase determination was performed using the dysprosium-derivative MAD data set and the program *SOLVE*/*RESOLVE* (Terwilliger & Berendzen, 1999) at 2.1 Å resolution (Table 1). The initial model was refined at 2.0 Å resolution by several iterations of refinement using *CNS* v.1.1 (Brünger *et al.*, 1998) and manual model revision using *QUANTA* (Accelrys Inc.) to yield R and R_{free} values of 0.21 and 0.24, respectively. The highest resolution limit was then extended to 1.6 Å by replacing the amplitude data with the native data set. The same refinement procedure was used to obtain the final model with R and R_{free} values of 0.197 and 0.212, respectively. The refinement statistics are summarized in Table 1.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outermost shell.

	Dy derivative			
	Edge	Peak	Remote	Native
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Wavelength (Å)	1.590996	1.591394	1.594101	1.54178
No. of reflections	179237	184475	199229	364118
Observed unique reflections	29969	30005	30005	59382
Resolution range	40.0–2.00 (2.07–2.0)	40.0–2.00 (2.07–2.0)	40.0–2.00 (2.07–2.0)	40.0–1.6 (1.66–1.60)
Completeness (%)	96.5	96.6	96.5	98.2
$I/\sigma(I)$	17.6 (4.9)	17.2 (5.0)	16.5 (5.0)	19.1 (6.2)
R_{merge}† (%)	7.1 (28.2)	9.2 (29.4)	6.2 (27.4)	4.3 (29.6)
Mean FOM (MAD/DM)	0.56/0.69	—	—	—
Refinement statistics				
No. of reflections observed	—	—	—	59331
Resolution range	—	—	—	37.0–1.6 (1.70–1.60)
R_{work}/R_{free} (%)	—	—	—	19.7 (23.4)/21.2 (27.0)
No. of molecules in ASU	—	—	—	2
No. of protein atoms	—	—	—	3600
No. of solvent molecules	—	—	—	568
R.m.s.d.				
Bond lengths (Å)	—	—	—	0.005
Angles (°)	—	—	—	1.2
Mean B factor (Å^2)				
All atoms	—	—	—	22.6
Protein atoms	—	—	—	20.4
Waters	—	—	—	36.4
Ramachandran plot				
Favoured	—	—	—	91.1
Allowed	—	—	—	8.4
Generous and disallowed	—	—	—	0.5

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of the the unique reflection hkl and $I_i(hkl)$ is the intensity of its ith observation.

2.6. Structure evaluation

The value of the dissociation energy ΔG for hydrophobic interactions was calculated based on comparison of the accessible surface area (ASA) of nonpolar and polar atoms in the native and denatured states (Funahashi *et al.*, 1999; Tanaka *et al.*, 2001). The ASA in the native state was calculated following the procedure of Connolly (1993). The ASA in the denatured state was calculated from an extended structure generated by *InsightII* (Accelrys Inc.). The entropic effect upon unfolding was estimated based on the amino-acid composition and the thermodynamic parameters reported previously (Oobatake & Ooi, 1993). The gap index was calculated as the ratio of volume and buried ASA for the dimer interface, which reflects the shape complementarity between interprotomer interfaces of homodimers (Jones & Thornton, 1996). These values were calculated with PROTORP (http://www.bioinformatics.sussex.ac.uk/protorp/).

2.7. Phosphatase activity measurement

The 1000 µl reaction mixture for the measurement of phosphatase activity consisted of 5 mM 4-nitrophenyl phosphate (4-NPP), 1 mM magnesium chloride, 0.2 M NaCl, 20 mM Tris–HCl pH 8.0 and 1 µg ml^{-1} purified PH1421 protein. The reaction was started by the addition of protein, which was followed by a 10 min incubation at 343 K. The reaction was stopped by the addition of 100 µl 1 N NaOH. The reaction mixture was centrifuged at 13 800g for 15 min at 277 K. The amount of the product 4-nitrophenol (4-NP) in the supernatant was determined photometrically at 405 nm using a molar absorption coefficient of 18 400 M^{-1} cm^{-1}. The activity is defined as the amount of 4-NP produced by one milligram of this enzyme per minute at 343 K.

3. Results and discussion

3.1. Amino-acid sequence analysis with databases

As the first step to obtain information about the structure–function relationship of PH1421, an amino-acid sequence analysis using the available databases was performed. A *PSI-BLAST* search revealed that most of the top 50 homologues were annotated as putative hydrolases/phosphatases belonging to the HAD superfamily. These proteins share full-length sequence identity of 24–38% with PH1421. As a next step, *PSI-BLAST* analysis against the PDB was performed to find PH1421 homologues with known three-dimensional structures. The search identified six hydrolases, TA0175 (PDB code 1l6r), YbiV (1rlo), YidA (1rkq), TM0651 (1nf2), BT4131 (1ymq) and phosphoserine phosphatase (PSP; 1f5s). All six enzymes are members of the HAD superfamily and their domain organizations indicate that they belong to subfamily II or the Cof-subfamily of this enzyme superfamily, with the exception of PSP. Indeed, the amino-acid sequence alignment confirmed the existence of three HAD signatures in the PH1421 sequence (Fig. 1). Collectively, these amino-acid

sequence analyses suggest that PH1421 is a member of the HAD superfamily. However, further investigation is necessary to obtain insight into the functional aspects of PH1421.

3.2. Overall structural description of the PH1421 protomer

To obtain further insight into the structure–function relationship, the crystal structure of PH1421 was determined by the multiple anomalous dispersion (MAD) method (Hendrickson *et al.*, 1990) using dysprosium(III) ion as an anomalous scatterer. The structure was refined at 1.6 Å resolution against the native data set (Table 1). The asymmetric unit of the PH1421 crystals contained a homodimer. The monomeric protomer can be divided into two distinct domains: a larger core domain and a smaller cap domain (Fig. 2*a*). The core domain is composed of two segments: residues 1–79 (β1-α1-β2-α2-β3-β4-β5) and residues 153–231 (α5-β9-α6-β10-α7-β11-α8). The spatial arrangement of the secondary structures is quite similar to that of the Rossmann fold and is well conserved among HAD-superfamily members. The cap domain (residues 80–152; α3-β6-α4-β7-β8) is topologically inserted between the two halves (residues 1–79 and 153–231) of the core domain (Fig. 2*c*). This insertion position of the cap domain suggests that PH1421 belongs to subfamily II or the Cof-subfamily.

3.3. PH1421 dimer in the asymmetric unit

In the crystallographic asymmetric unit, PH1421 has been found to be homodimeric (Fig. 2*b*). The two protomers in the dimer are related to each other by a noncrystallographic pseudo-twofold axis. The dimer interfaces can be classified into two parts: an interprotomer hydrophobic core including well conserved nonpolar residues between the two cognate core domains and surrounding interprotomer contacts consisting of mainly polar residues (Figs. 3*b* and 4). The latter contacts include two hydrogen bonds (two pairs of Gln48–Asp108 interactions) and four charged hydrogen bonds (two pairs of Arg77–Asp108 interactions).

The buried ASA on dimerization has been calculated as 1250 Å^2 per protomer or 11% of the total surface ASA of the protomer (Table 2), suggesting that the dimer association is strong enough for an oligomeric protein (Janin, 1997). The gap index is one of the parameters used to evaluate the shape complementarity of interprotomer interfaces; it is defined as

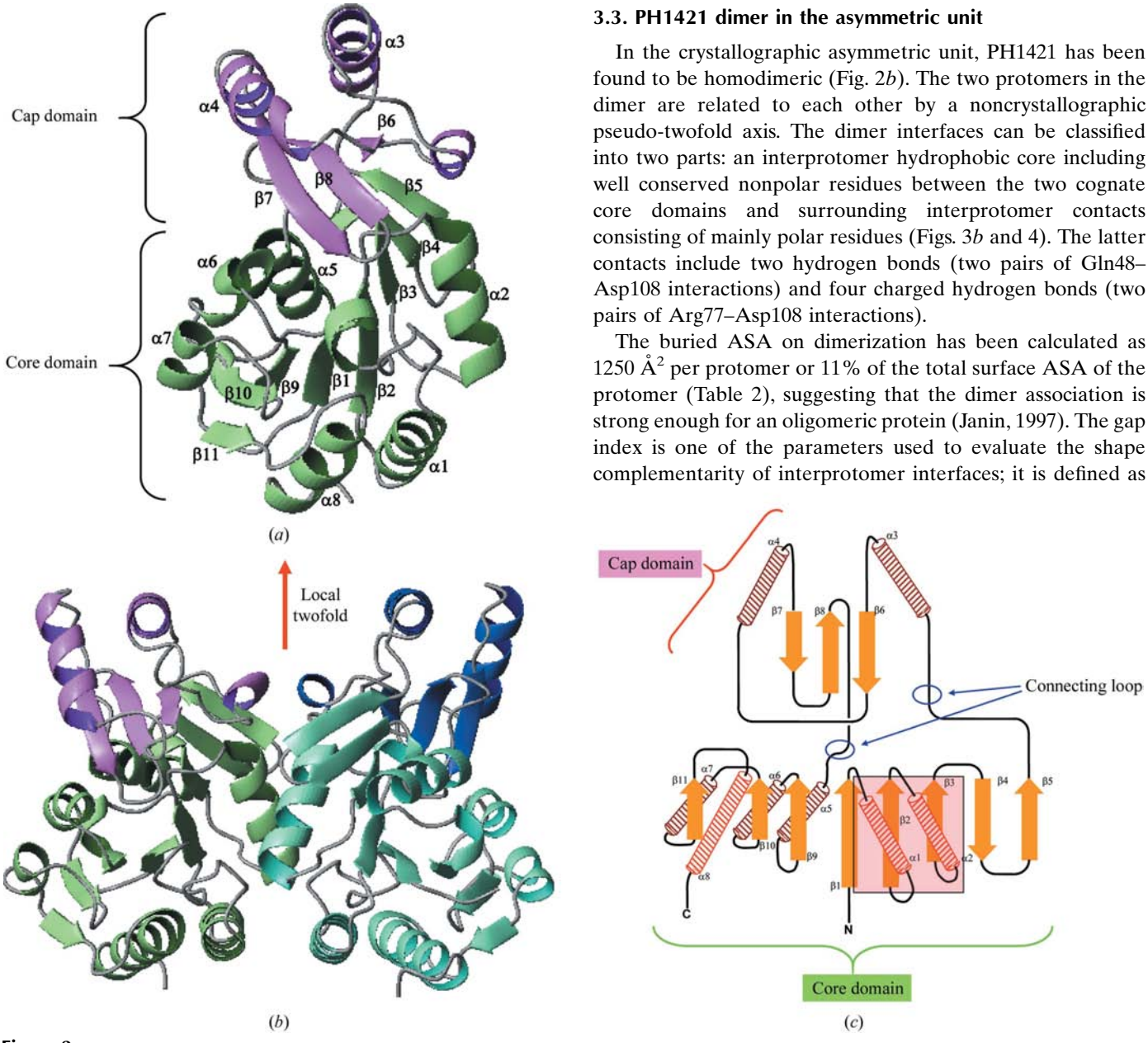


Figure 2
The crystal structure of PH1421. (*a*) Ribbon representation of the protomer. The two domains are distinguished by colour. α-Helices and β-strands are labelled individually. (*b*) Ribbon representation of the dimer. The four domains are distinguished by colour. A noncrystallographic pseudo-twofold axis which defines the spatial relationship between the two subunits is shown. These two figures were produced using *MOLMOL* (Koradi *et al.*, 1996). (*c*) Topology diagram of the secondary structure. Cylinders and broad arrows denote α-helices and β-strands, respectively. Note that α1 and α2 (highlighted by a red shaded box) are involved in the dimer interface.

the ratio of the volume and buried ASA for the dimer interface and roughly expresses the mean distance between the interfaces of the two subunits (Jones & Thornton, 1996). The gap index of PH1421 has been calculated to be 2.72 Å. This is consistent with the typical value of 2.2 ± 0.9 Å for a homodimeric interface. Any possible crystallographic symmetry operation cannot produce other dimers with effective buried ASA values. These results suggest that the dimer association in the crystallographic asymmetric unit is the biologically relevant form in solution; this is further supported by the dynamic light-scattering experiment results, which show a dimeric state of PH1421 in solution (see §2.2).

3.4. Enzymatic activity of PH1421

The crystal structure of PH1421 revealed a putative active site surrounded by the well conserved HAD motifs I, II and III (Figs. 1, 5*a* and 5*c*). These motifs spatially gather to form a putative active-site pocket between the cap and core domains. These features suggest that PH1421 probably hydrolyzes phosphorylated substrate as a HAD-superfamily enzyme. To confirm the phosphatase activity of PH1421, the hydrolysis of 4-nitrophenyl phosphate (4-NPP), a general substrate widely used to assess phosphatase activity, was examined in the presence or absence of Mg^{2+} ion. In the presence of Mg^{2+}, a 4-NPP hydrolase activity of 239.1 nmol min^{-1} mg^{-1} was observed. In contrast, in the absence of Mg^{2+} release of the product 4-NP was not detected. These results suggest that PH1421 is a magnesium-dependent phosphatase belonging to the HAD superfamily.

3.5. Comparison with TA0175

3.5.1. Protomer-fold similarity. At this point, as the three-dimensional information on PH1421 became available, the PDB was screened using the *DALI* algorithm (Holm &

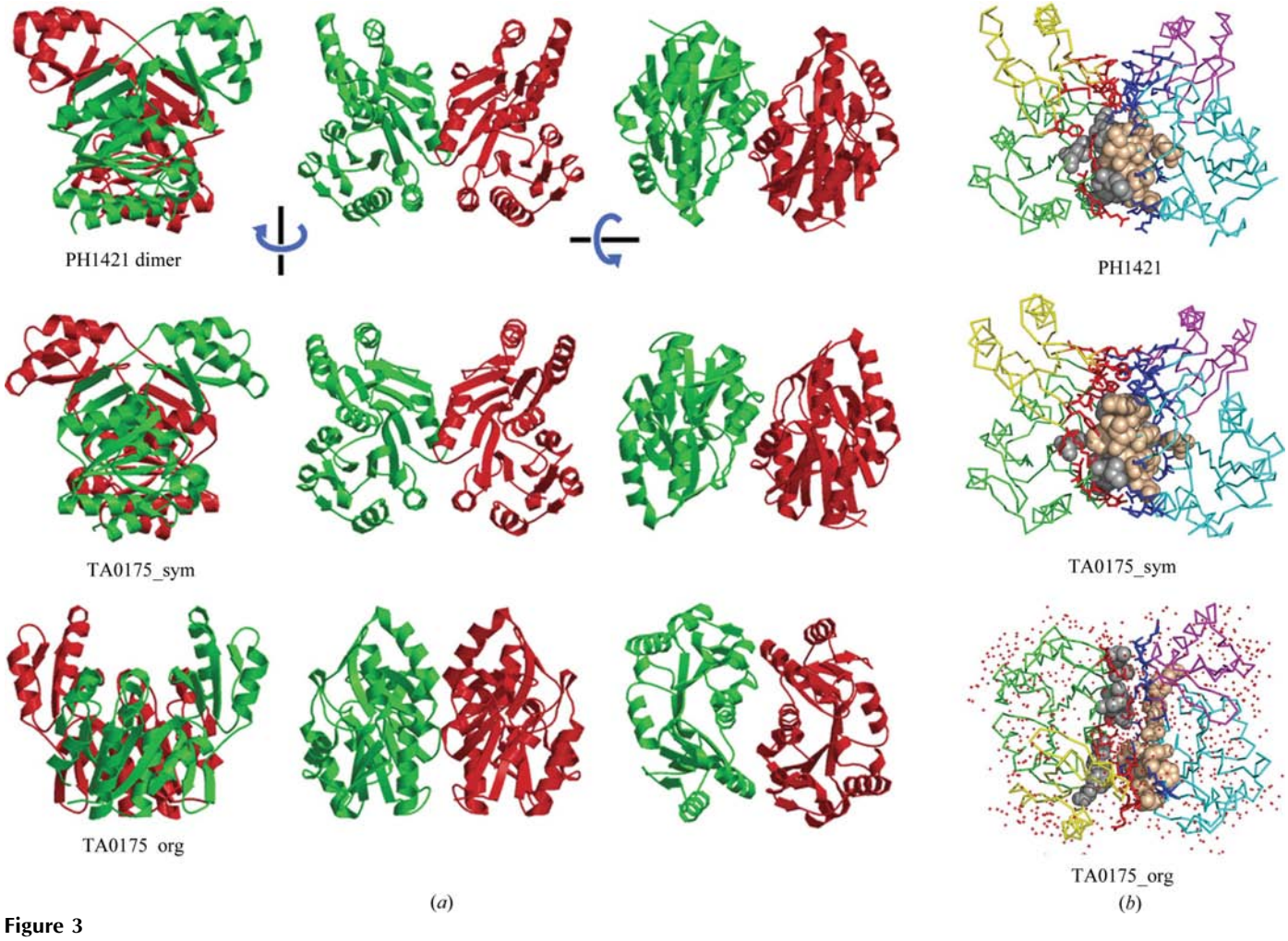


Figure 3
The dimeric states of PH1421 and TA0175. (*a*) Ribbon representations of the dimer from three orthogonal views. The two subunits in the dimer are distinguished by colour. Top, the PH1421 dimer in this study. Middle, a TA0175 dimer generated by a crystallographic symmetry operation, abbreviated as TA0175_sym. Bottom, the TA0175 dimer reported previously, abbreviated as TA0175_org (Kim *et al.*, 2004). This figure was produced using *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997). (*b*) Dimer-interface residues. Models are represented as C^{α} traces. One protomer (MolA) is coloured yellow for the cap domain and green for the core domain; the other (MolB) is coloured violet for the cap domain and cyan for the core domain. Space-filling models represent the interprotomer hydrophobic core consisting of nonpolar residues; the subunit affiliation is distinguished by colour. Stick models represent the interface residues surrounding the hydrophobic core; the subunit affiliation is distinguished by colour. In the TA0175_org case, water molecules are shown as red dots. This figure was produced using *PyMOL* (DeLano, 2002).

Sander, 1997) to find structural homologues. In agreement with the results of amino-acid sequence analysis, numerous HAD enzymes were found as structural homologues. Of these structures, TA0175 (PDB code 1l6r or 1kyt; 28% sequence identity with PH1421) was listed with the highest Z score of 26.6 (the second highest was 15.3). Further structural comparison confirmed the high structural similarity between PH1421 and TA0175. The corresponding C^{α} atoms of the two structures are superimposable with a root-mean-square deviation (r.m.s.d.) of 2.1 Å (Fig. 5*b*). In the core domains, the spatial arrangements of the HAD-motif residues are well conserved in the putative active sites of both enzymes (Figs. 5*a* and 5*c*). Furthermore, these two enzymes are also quite similar in the overall folding of the cap domain (Figs. 5*a* and 5*b*). Apart from TA0175, no other structure with significant structural similarity in the cap domain was found by the *DALI* analysis. The substrate-specificity loop (residues 136–145) of TA0175 corresponds to residues 139–148 of PH1421 (Figs. 5*a* and 5*b*). In these loops, Ser138, Ser141 and His143 in TA0175, which correspond to Ser141, Ala144 and His146 in PH1421, respectively, are thought to be involved in substrate recognition (Lu *et al.*, 2005). Taken together with the fact that the cap domain defines substrate recognition, it is suggested that PH1421 and TA0175 share closely related biological roles or catalytic functions.

3.5.2. Interprotomer association. Although the HAD enzymes PH1421 and TA0175 are suggested to be closely related, a striking difference between their crystal structures has been found. Both PH1421 and TA0175 are dimeric in solution, which was confirmed by a dynamic light-scattering experiment for PH1421 and by gel-filtration analysis and a kinetic experiment for TA0175 (Kim *et al.*, 2004). However, the proposed models of the biological dimer based on the crystal structures are completely different for these two enzymes (Fig. 3). In the case of PH1421, only one biologically relevant dimer association is observed in the crystal structure (Fig. 3, top). In contrast, for the crystal structure of TA0175, in addition to the dimer association reported previously (denoted as

Table 2
Factors contributing to the thermal stability of PH1421 and TA0175.

ΔG and $\Delta\Delta G$ values are in kJ mol^{-1}.

(*a*) Interprotomer.

		Accessible surface area (ASA)										
		Total ASA		Buried ASA		Hydrophobicity		Hydrogen bond (<3.4 Å)				
Protein†	Residue	Monomer (Å^2)	Dimer (Å^2)	Monomer (Å^2)	(%)	ΔG	$\Delta\Delta G$	No.	ΔG	$\Delta\Delta G$	Ion pair (<5 Å)	Gap index (Å)
PH1421 (1wr8)	231	10702	18902	1250	11.1	147	—	9	75	—	3	2.72
TA0175_org (1l6r)	224	10331	19289	687	6.7	53	−94	3	25	−50	2	4.85
TA0175_sym‡	—	10331	18291	1185	12.1	147	0	8	66	−9	4	2.18

(*b*) Intraprotomer.

		Hydrophobicity		Hydrogen bond (<3.4 Å)				Entropy effect of amino-acid side chain		
Protein†	Residue	ΔG	$\Delta\Delta G$	No.	ΔG	$\Delta\Delta G$	Ion pair (<5 Å)	ΔG	$\Delta\Delta G$	Proline content (%)
PH1421 (1wr8)	231	2557	—	253	2100	—	18	−117	—	3.5
TA0175 (1l6r)	224	2490	−67	268	2224	124	13	−252	−135	2.6

† PDB codes used for calculations are given in parentheses. ‡ A dimer generated by a crystallographic symmetry operation.

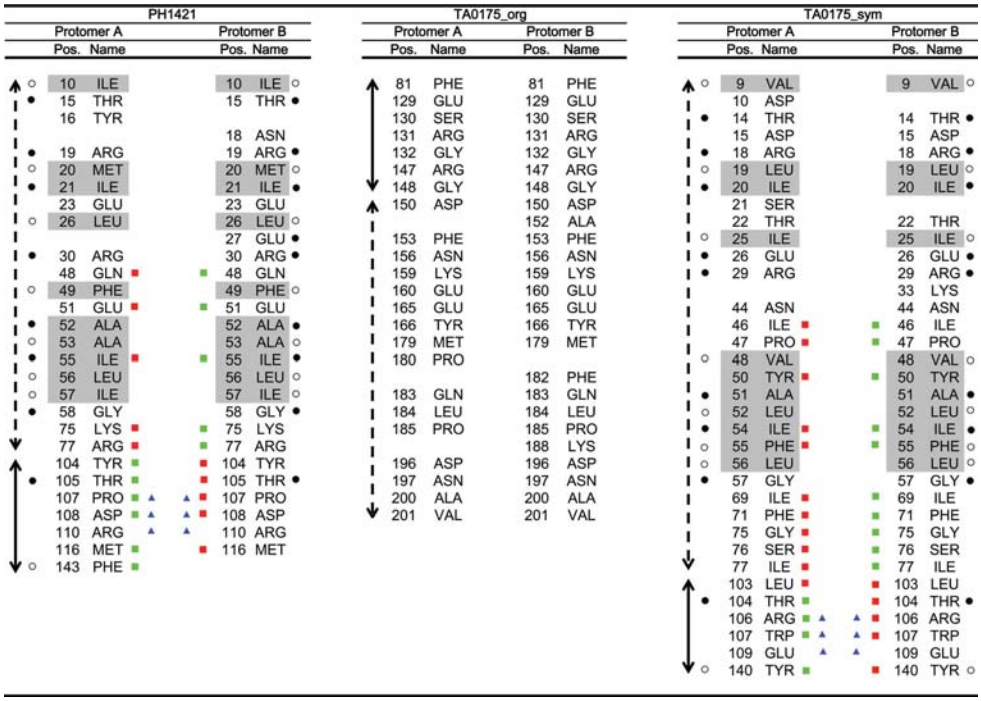
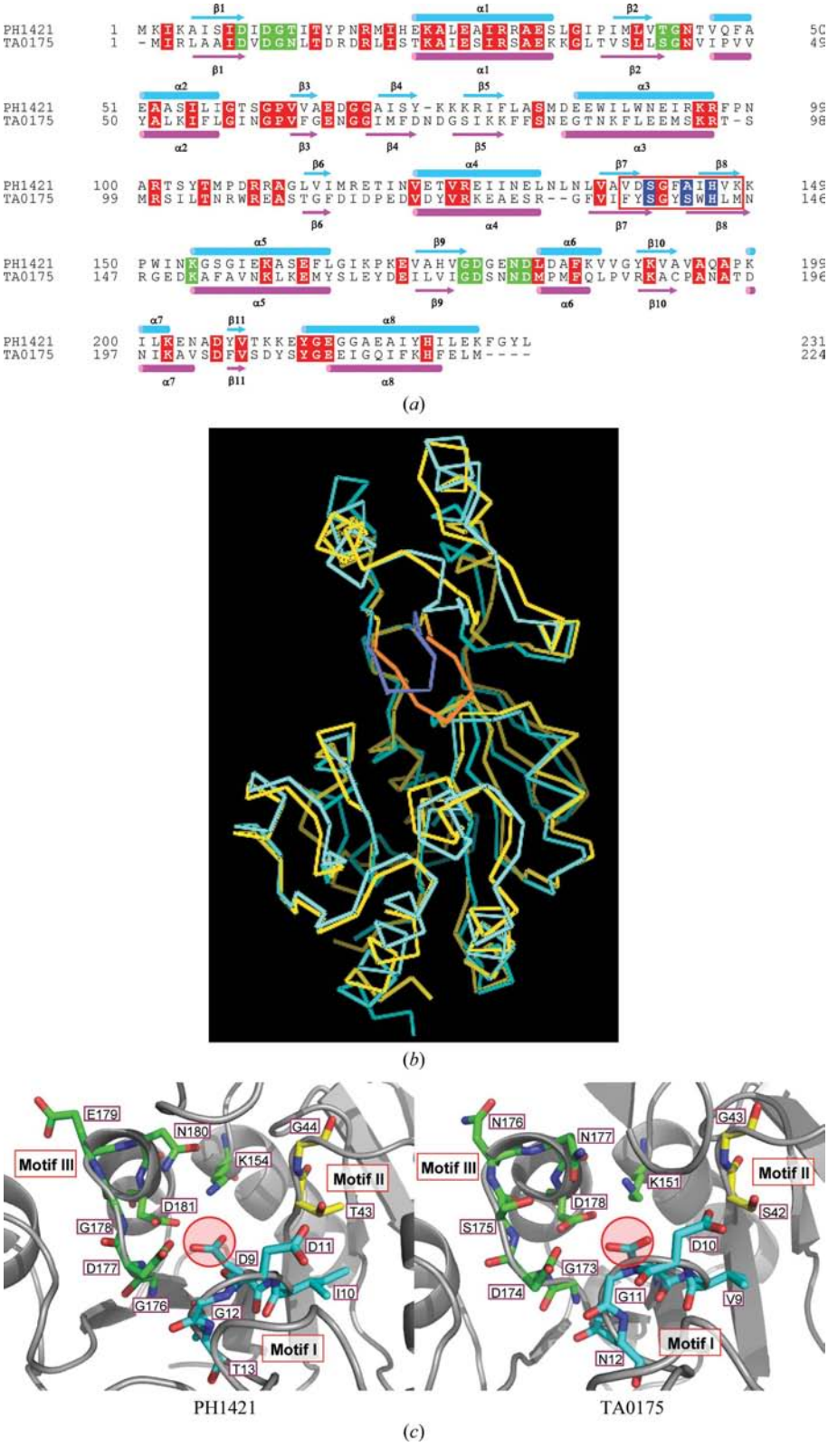


Figure 4
Interface residues of the two HAD phosphatases. Broken and solid arrows indicate core-domain and cap-domain residues, respectively. Shaded letters indicate nonpolar residues in the inter-protomer hydrophobic cores of the PH1421 and TA0175_sym dimers. Black and white circles indicate residues that are identical and similar in the PH1421 and TA0175_sym dimers, respectively. Squares and triangles indicate cap–core and cap–cap interface residues, respectively; residues with the same colour interact with each other.

TA0175_org; Kim *et al.*, 2004; PDB codes 1l6r and 1kyt), another dimer association (TA0175_sym) can be generated by a crystallographic symmetry operation which is obviously quite similar to that of the PH1421 dimer (Fig. 3, bottom and middle). In order to evaluate the biological relevance of the two dimers, the interprotomer interactions were investigated by calculating several thermal parameters. The results are summarized in Table 2 for TA0175_org, TA0175_sym and PH1421: (i) the interprotomer buried ASAs are 687, 1185 and 1250 Å^2, respectively; (ii) the Gibbs energy differences (ΔG) arising from hydrophobic interactions on the dimer interface are 53, 147 and 147 kJ mol^{-1}, respectively; (iii) the ΔG values arising from interprotomer hydrogen bonds are 25, 66 and 75 kJ mol^{-1}, respectively; (iv) the numbers of interprotomer ion pairs are two, four and three, respectively; and (v) the gap indices are 4.85, 2.18 and 2.72 Å, respectively. Clearly, all five parameters for TA0175_sym are quite similar or identical to those for PH1421, while the parameters for TA0175_org differ from those for TA0175_sym and PH1421. These results are consistent with the results of visual inspection of the protomer interfaces of the two TA0175 dimer forms. The interprotomer space in TA0175_org seems to be filled with intervening water molecules (Fig. 3*b*, bottom), while that of TA0175_sym forms a hydrophobic core with well conserved nonpolar residues (Figs 3*b* middle and 4). This fact suggests that the association of TA0175_sym is energetically more favourable and stable than that of TA0175_org. Obviously, the TA0175_sym dimer interface is also



(*a*)

(*b*)

(*c*)

Figure 5
Comparison between PH1421 and TA0175. (*a*) Structure-based amino-acid sequence alignment. The secondary-structure assignments by *DSSP* (Kabsch & Sander, 1983) for PH1421 and TA0175 are shown above and below the aligned sequences, respectively. Conserved active-site residues for phosphatase activity are highlighted in green. The sequences of the putative specificity loops are boxed. The putative residues for substrate recognition are highlighted in blue. Other conserved residues are highlighted in red. (*b*) C^{α} superposition of PH1421 and TA0175 protomers. PH1421 and TA0175 are coloured cyan and yellow, respectively. The putative substrate-specificity loops of PH1421 (residues 139–148) and TA0175 (residues 136–145) are coloured purple and orange, respectively. This figure was produced using *PyMOL* (DeLano, 2002). (*c*) Conserved active-site residues for phosphatase activity. Three conserved HAD motifs are indicated. Aspartic acids acting as nucleophiles are enclosed in red circles. The consensus residues in the three HAD motifs are labelled individually.

formed in a similar manner to that of PH1421 through an inter-protomer hydrophobic core consisting of nonpolar residues between the two cognate core domains and surrounding inter-protomer contacts consisting of polar residues (Fig. 3*b*, top and middle; Fig. 4). Taken together, it would be reasonable to conclude that TA0175_sym dimer association is the biologically relevant form of TA0175 in solution.

Integrating the results described in the preceding sections, PH1421 could be tentatively characterized as a HAD phosphatase that shares similar structural features and biological functions with TA0175. The conservation of dimeric association in the two enzymes implies some biological significance of the dimeric state. One possible role is a structural basis for the positive cooperativity in catalysis. The TA0175 dimer has been reported to have a positive cooperativity for inorganic diphosphate with a Hill's coefficient n of 1.7 ± 0.2 (Kim *et al.*, 2004). Such catalytic behaviour would require intimate 'cross-talk' between the two protomers so that ligand binding to the active site on one protomer modifies the affinity of the other active site on the cognate protomer. Similar tight dimeric organization of PH1421 and TA0175 may be essential for communication of binding information between the active sites of the dimers. Another possible role is a thermoadaptation to ensure effective catalysis of these two thermophilic HAD enzymes at high temperature. It has been reported that the cap and core domains perform open–closed movement upon ligand binding for the desolvation of the active site from the bulk solvent (Lu *et al.*, 2005; Burroughs *et al.*, 2006). Since the cap and core domains are linked by only two flexible connecting loops (Fig. 2*c*), such an unstable connection would require some structural restraints against thermal disturbance. In fact, the thermophilic enzyme TA0175 shows high catalytic efficiency and substrate specificity (K_m and k_{cat}/K_m of 0.037 ± 0.003 mM and $2.2 \times 10^5\ M^{-1}\ s^{-1}$ for 2-phosphoglycolate and of 3.1 ± 0.5 mM and $0.19 \times 10^3\ M^{-1}\ s^{-1}$ for 4-NPP, respectively) at 343 K (Kim *et al.*, 2004). The mobile cap domain seems to be fixed in the same way in both dimers (Fig. 4): (i) the interprotomer interface loop (residues 104–110 of PH1421 and residues 103–109 of TA0175) in the cap domain of one subunit interacts with the core-domain residues involving helix $\alpha2$ and strand $\beta5$ on the other subunit (cap–core interactions) and (ii) the same loop also interacts with the cognate cap domain (cap–cap interactions). The mode of dimeric association shown in this study may explain such structural restraint in the two HAD enzymes. HAD enzymes have been known to have low substrate specificity in general. Tremblay and coworkers discussed the possibility that the HAD activity is controlled by cellular compartmentalization or gene expression (Tremblay *et al.*, 2006). Further investigation is required to examine this possibility.

3.5.3. Thermostability. Since PH1421 from *P. horikoshii* OT3 (living temperature 371 K) has been found to be a close structural homologue of TA0175 from the mesothermophilic *T. acidophilum* (living temperature 338 K), comparison of these two structures should provide insight into the thermostabilization mechanism of this enzyme. The stabilizing factors for the protomer were investigated by a semi-empirical evaluation of thermodynamic parameters (Table 2). Hydrophobic interactions in the interior of the protein are an important stabilizing factor (Kauzmann, 1959). The ΔG from hydrophobic interaction was calculated as 2557 and 2490 kJ mol^{-1} for PH1421 and TA0175, respectively. This result indicates that hydrophobic interactions may play a certain role in the stabilization of the PH1421 protomer. Electrostatic interactions represent a significant stabilizing force in the folded protein (Hennig *et al.*, 1995; Yip *et al.*, 1995; Pappenberger *et al.*, 1997). The number of intraprotomer ion pairs (<5 Å) for PH1421 and TA0175 has been calculated as 18 and 13, respectively, suggesting that ionic interactions contribute significantly to the higher stability of the PH1421 protomer. Decreasing the entropy of the unfolded state can stabilize proteins (Matthews *et al.*, 1987). This effect is achieved by selecting amino-acid residues that have lower entropy in the unfolded state. This effect arising from amino-acid side chains can be calculated from the amino-acid composition (Oobatake & Ooi, 1993). The ΔG arising from the entropic effect upon unfolding of PH1421 and TA0175 is estimated as -117 and -252 kJ mol^{-1}, respectively. The percentage of proline residues has been found to be 3.5% and 2.6% for PH1421 and TA0175, respectively. These results suggest that the entropic effect contributes significantly to the higher stability of the PH1421 protomer and that the higher proline content dominates this effect. In contrast, the ΔG arising from intraprotomer hydrogen bonds has been calculated as 2100 and 2224 kJ mol^{-1} for PH1421 and TA0175, respectively. This result suggests that hydrogen bonds do not contribute to the higher stability of the PH1421 protomer. Generally, oligomerization has been recognized as one of the thermostabilization factors of proteins. As discussed in §3.5.2, the dimer associations of PH1421 and TA0175 are quite similar. However, the thermostabilization effect of oligomerization should be investigated further, especially using a mesophilic orthologue.

To summarize, the higher thermal stability of the hyperthermophilic PH1421 when compared with the mesothermophilic TA0175 is achieved by higher stability of the protomer. The potential contributing factors are (i) an increase in hydrophobic interactions, (ii) an increase in ion pairs and (iii) an entropic effect from the amino-acid composition.

4. Conclusions

PH1421 is equipped with several features typical of HAD-family phosphatases: (i) a conserved HAD-domain organization and core-domain fold, (ii) conserved active-site residues and (iii) magnesium-dependent phosphatase activity. In particular, PH1421 shares the following similar structural features with the dimeric HAD phosphatase TA0175 from *T. acidophilum* harbouring 2-glycolate phosphatase activity: (i) the fold and secondary-structure arrangement of the cap domain, (ii) the cap-insertion position in the core domain, which is indicative of HAD subfamily II enzymes, and (iii) dimeric association with a well conserved tight hydrophobic core. These results suggest that PH1421 is a magnesium-

dependent HAD phosphatase that acts on compounds closely related to 2-phosphoglycolate. The common dimeric structure of PH1421 and TA0175 may be important for their enzymatic function. The higher thermostability of PH1421 can be attributed to thermostabilization of the protomer arising from hydrophobic interactions, ion pairs and the entropic effect.

HY performed the crystallization, structure determination and enzymatic analysis and prepared the paper, KT contributed to the large-scale protein production, MS contributed to the automatic crystallization screening and NK supervised the work. The authors would like to thank the staff of RIKEN Genomic Science Center for providing the plasmid and the technical staff of RIKEN SPring-8 Center for assistance with the experiments. We also thank M. Yamamoto and his staff for assistance during data collection on beamline BL26B1 of SPring-8. This work (PH1421/HTPF10283) was supported by the 'National Project on Protein Structural and Functional Analyses' funded by the MEXT of Japan.

References

Allen, K. N. & Dunaway-Mariano, D. (2004). *Trends Biochem. Sci.* **29**, 495–503.

Aravind, L., Galperin, M. Y. & Koonin, E. V. (1998). *Trends Biochem. Sci.* **23**, 127–129.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D. & Aravind, L. (2006). *J. Mol. Biol.* **361**, 1003–1034.

Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J, Higgins, D. G. & Thompson, J. D. (2003). *Nucleic Acids Res.* **31**, 3497–3500.

Collet, J. F., Stroobant, V., Pirard, M., Delpierre, G. & Van Schaftingen, E. A. (1998). *J. Biol. Chem.* **273**, 14107–14112.

Connolly, M. L. (1993). *J. Mol. Graph.* **2**, 139–141.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://pymol.sourceforge.net.

Funahashi, J., Takano, K., Yamagata, Y. & Yutani, K. (1999). *Protein Eng.* **12**, 841–850.

Galburt, E. A., Pelletier, J., Wilson, G. & Stoddard, B. L. (2002). *Structure*, **10**, 1151–1152.

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.

Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J. N. (1995). *Structure*, **3**, 1295–1306.

Holm, L. & Sander, C. (1997). *Nucleic Acids Res.* **25**, 231–234.

Janin, J. (1997). *Nature Struct. Biol.* **4**, 973–974.

Jones, S. & Thornton, J. M. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 13–20.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Kauzmann, W. (1959). *Adv. Protein Chem.* **14**, 1–63.

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.

Koradi, R., Billeter, M. & Wuthrich, K. (1996). *J. Mol. Graph.* **14**, 51–55.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Lahiri, S. D., Zhang, G., Dai, J., Dunaway-Mariano, D. & Allen, K. N. (2004). *Biochemistry*, **43**, 2812–2820.

Lahiri, S. D., Zhang, G., Dunaway-Mariano, D. & Allen, K. N. (2002). *Biochemistry*, **41**, 8351–8359.

Lahiri, S. D., Zhang, G., Dunaway-Mariano, D. & Allen, K. N. (2003). *Science*, **299**, 2067–2071.

Lu, Z., Dunaway-Mariano, D. & Allen, K. N. (2005). *Biochemistry*, **44**, 8684–8696.

Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 6663–6667.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D. & Allen, K. N. (2000). *Biochemistry*, **39**, 10385–10396.

Oobatake, M. & Ooi, T. (1993). *Prog. Biophys. Mol. Biol.* **59**, 237–284.

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

Pappenberger, G., Schuring, H. & Jaenicke, R. (1997). *J. Mol. Biol.* **274**, 676–683.

Peisach, E., Selengut, J. D., Dunaway-Mariano, D. & Allen, K. N. (2004). *Biochemistry*, **43**, 12770–12779.

Roberts, A., Lee, S.-Y., McCullagh, E., Silversmith, R. E. & Wemmer, D. E. (2005). *Proteins*, **58**, 790–801.

Selengut, J. D. (2001). *Biochemistry*, **40**, 12704–12711.

Shin, D. H., Roberts, A., Jancarik, J., Yokota, H., Kim, R., Wemmer, D. E. & Kim, S.-H. (2003). *Protein Sci.* **12**, 1464–1472.

Sugahara, M. & Miyano, M. (2002). *Tampakushitu Kakusan Koso*, **47**, 1026–1032.

Tanaka, H., Chinami, M., Mizushima, T., Ogasahara, K., Ota, M., Tsukihara, T. & Yutani, K. (2001). *J. Biochem.* **130**, 107–118.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Tremblay, L. W., Dunaway-Mariano, D. & Allen, K. N. (2006). *Biochemistry*, **45**, 1183–1193.

Ueno, G., Kanda, H., Hirose, R., Ida, K., Kumasaka, T. & Yamamoto, M. (2006). *J. Struct. Funct. Genomics*, **1**, 15–22.

Wang, W., Kim, R., Jancarik, J., Yokota, H. & Kim, S.-H. (2001). *Structure*, **9**, 65–71.

Yip, K. S., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. & Rice, D. W. (1995). *Structure*, **3**, 1147–1158.