Crystal Structure of the Hyperthermophilic Inorganic Pyrophosphatase from the Archaeon *Pyrococcus horikoshii*

Binbin Liu,*‡ Mark Bartlam,*‡ Renjun Gao,† Weihong Zhou,* Hai Pang,* Yiwei Liu,* Yan Feng,† and Zihe Rao*‡
*Laboratory of Structural Biology, Tsinghua University and National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China; †Key Laboratory for Molecular Enzymology and Engineering of Ministry of Education, Jilin University, Changchun 130023, China; and ‡National Laboratory of Macromolecules, Institute of Biophysics, Chinese Academy of Science, Beijing 100101, China

ABSTRACT A homolog to the eubacteria inorganic pyrophosphatase (PPase, EC 3.6.1.1) was found in the genome of the hyperthermophilic archaeon *Pyrococcus horikoshii*. This inorganic pyrophosphatase (Pho-PPase) grows optimally at 88°C. To understand the structural basis for the thermostability of Pho-PPase, we have determined the crystal structure to 2.66 Å resolution. The crystallographic asymmetric unit contains three monomers related by approximate threefold symmetry, and a hexamer is built up by twofold crystallographic symmetry. The main-chain fold of Pho-PPase is almost identical to that of the known crystal structure of the model from *Sulfolobus acidocaldarius*. A detailed comparison of the crystal structure of Pho-PPase with related structures from *S. acidocaldarius*, *Thermus thermophilus*, and *Escherichia coli* shows significant differences that may account for the difference in their thermostabilities. A reduction in thermolabile residues, additional aromatic residues, and more intimate association between subunits all contribute to the larger thermophilicity of Pho-PPase. In particular, deletions in two loops surrounding the active site help to stabilize its conformation, while ion-pair networks unique to Pho-PPase are located in the active site and near the C-terminus. The identification of structural features that make PPases more adaptable to extreme temperature should prove helpful for future biotechnology applications.

INTRODUCTION

Inorganic pyrophosphatase (PPase, EC 3.6.1.1) is an essential enzyme that specifically catalyzes the hydrolysis of the phosphoanhydride bond in inorganic phosphate (PPi) (Chen et al., 1990; Lahti, 1983). PPi is a central phosphorus metabolite and is a by-product of various reversible nucleoside 58-triphosphate–dependent reactions including tRNA charging and DNA and protein synthesis that utilize ATP in vivo. PPases from a wide variety of sources have been studied, but those from *Saccharomyces cerevisiae* and *Escherichia coli* are the most highly characterized, both biochemically and structurally. *E. coli* pyrophosphatase is a homohexamer with 175 amino acids per monomer (Avaeva et al., 1997), whereas *S. cerevisiae* pyrophosphatase is a homodimer with 286 amino acids per monomer (Heikinheimo et al., 2001).

PPases from archaebacterium exhibit different structural and catalytic properties (Hansen et al., 1999; Richter and Schafer, 1992). The archaeal PPases so far reported are relatively thermostable, especially in the presence of divalent metal cations (Ichiba et al., 1998). Understanding the structural basis for the enhanced stability of proteins from hyperthermophilic organisms relative to their mesophilic and thermophilic counterparts is a highly relevant but complex and challenging problem. Previous comparisons of high-resolution crystal structures of enzymes with the same fold and function in mesophiles, thermophiles and hyperthermophiles have revealed a number of potentially stabilizing features. A proper understanding of the molecular basis of thermal stability in proteins could have important consequences for their application in a range of biotechnological processes. For example, thermostable pyrophosphatases have common uses in cycle sequencing methods using thermostable DNA polymerases (Vander Horn et al., 1997). The crystal structures of PPases from thermophilic bacterium *Thermus thermophilus* (T-PPase; PDB ID 2PRD) (Teplyakov et al., 1994), thermophilic archaebacterium *Sulfolobus acidocaldarius* (S-PPase; PDB ID 1QEZ) (Leppanen et al., 1999), and mesophile *E. coli* (E-PPase; PDB ID 1JFD) (Avaeva et al., 1997) have provided basic clues for the PPase catalytic mechanism and thermostability, but many important aspects remain to be resolved.

Microorganisms can be classified according to their optimal growth temperature, T_{opt}, into four groups: psychrophilic ($0 < T_{opt} < 20$°C), mesophilic ($20 < T_{opt} < 50$°C), thermophilic ($50 < T_{opt} < 80$°C), and hyperthermophilic ($80 < T_{opt} < 120$°C). Considerable efforts have been made during recent years to analyze the structural features that determine the extraordinary thermal stability of proteins from hyperthermophiles. Here we have isolated an inorganic pyrophosphatase (Pho-PPase) from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3, whose optimum growth temperature (95°C) is significantly higher than those of *S. acidocaldarius* (75–80°C), *T. thermophilus* (75–80°C) and *E. coli* (37°C). Pho-PPase showed a higher optimal activity at 88°C and an alkaline optimal pH of 10.3 (at 88°C). The enzyme has extreme thermostability and does not lose

Submitted June 12, 2003, and accepted for publication August 27, 2003.

Binbin Liu and Mark Bartlam contributed equally to this work.

Address reprint requests to Zihe Rao, Laboratory of Structural Biology, Dept. of Biological Science and Biotechnology, Tsinghua University, Beijing 100084, P. R. China. Tel.: 86-10-6277-1493; Fax: 86-10-6277-3145; E-mail: raozh@xtal.tsinghua.edu.cn.

0006-3495/04/01/420/08 $2.00

activity at 100°C. In addition, Pho-PPase is stable against various denaturants. All of these properties are different from those of other archaeal PPases: full details of the characterization of inorganic Pho-PPase will be reported elsewhere (Feng et al., unpublished results). To gain a more penetrating insight into its function, here we describe the structure determination of Pho-PPase and a comparison of the structure with its mesophilic and thermophilic counterparts in an attempt to understand the structural basis for thermal stability.

MATERIALS AND METHODS

Crystallization and x-ray data collection

The preparation and preliminary characterization of Pho-PPase crystals have been described elsewhere (B. Liu X. Li, R. Gao, W. Zhou, G. Xie, M. Bartlam, H. Pang, Y. Feng, and Z. Rao, submitted). Briefly, the gene encoding inorganic pyrophosphatase from the archaea *P. horikoshii* was cloned into pET15b (Novagen) and expressed in *E. coli* strain BL21. After purification, the purified protein was concentrated to 20 mg/ml. Crystallization trials were set up using the hanging drop/vapor diffusion method with Crystal Screen reagent kits I and II (Hampton Research). Crystals suitable for diffraction were obtained after two weeks from the condition 3.8% PEG 4000, 0.1 M Na acetate, pH 5.0–5.2, 0.02 M $MgCl_2$. A set of data at 2.66 Å resolution were collected in house. All data were integrated using DENZO/HKL and scaled and merged with SCALEPACK (Otwinowski and Minor, 1997). Crystal lattice properties and data-collection statistics are listed in Table 1.

Structure determination and refinement

The structure of Pho-PPase was solved by the molecular-replacement (MR) method using the protein S-PPase (PDB ID 1QEZ) (Leppanen et al., 1999) as the starting model. The initial *R* value for the MR solution obtained from the cross-rotation and translation search in CNS (Brunger et al., 1998) was 48.4%, using 15171 reflections in the 50-2.66 Å resolution range. This *R* value is ~10% lower than those of the other possible solutions. After rigid-body refinement, the *R* value decreased to 46.8, and side-chain atoms were fitted into the 2|Fo|-|Fc| electron-density map. The structure was further refined to 26.5% (using reflections in the resolution range 50-2.66 Å) following cycles of simulated-annealing refinement using CNS and manual rebuilding in O (Jones et al., 1991). After the placement and refinement of 93 water molecules and individual *B* factor refinement, the *R* factor was reduced to 23.2%. A Ramachandran plot generated by PROCHECK (Laskowski et al., 1993) shows that the structure has reasonable stereochemistry with no residues in disallowed regions. Refinement statistics are summarized in Table 1. Coordinates for this structure have been deposited in the Protein Data Bank with PDB ID 1UDE.

TABLE 1 Data collection, refinement, and model statistics

	Data collection
Space group	$P2_12_12$
Unit cell parameters	a = 71.8, b = 86.7, c = 92.8 Å $\alpha = \beta = \gamma = 90°$
Matthews coefficient ($Å^3$ Da^{-1})	1.93
Solvent content (%)	35.65
Resolution (Å)	50-2.66
Total observations	118041
Unique reflections	17030
Redundancy	6.93
Average $I/\sigma(I)$	12.1 (4.3)
R_{merge} (%)	7.5 (13.9)
Completeness (%)	99.8 (98.0)
	Refinement
Reflections	
(observed)	117986
(test)	1687
Resolution range (Å)	50-2.66
Protein atoms	4148
Solvent atoms	93
R_{work} (%)	23.24
R_{free} (%)	27.65
Average *B* value ($Å^2$)	38.19
RMSD bonds (Å)	0.009
RMSD angles (°)	1.6
Ramachandran plot (%)	
Most favored regions	76.9%
Additionally allowed regions	19.6%
Generously allowed regions	3.3%
Disallowed regions	0.2%

RESULTS AND DISCUSSION

Monomeric structure

The current model of *P. horikoshii* inorganic pyrophosphatase includes residues 4–171 in chain A, 4–170 in chain B, 4–170 in chain C, and 93 water molecules. Pho-PPase is arranged in a globular form and belongs to the $\alpha+\beta$ class of protein folds. The Pho-PPase monomer structure is composed of nine β-strands and two α-helices arranged in a $\beta1$-$\beta2$-$\beta3$-$\beta4$-$\beta5$-$\beta6$-$\beta7$-$\beta8$-$\alpha1$-$\beta9$-$\alpha2$ topology (Fig. 1). Pho-PPase shares 47% sequence identity with S-PPase (Fig. 2 *A*), and the two proteins have a similar core structure. Indeed, superposition of the Pho-PPase structure with S-PPase, T-PPase, and E-PPase shows that the four PPases are spatially homologous (Fig. 2 *B*). The RMSD between C_α atoms of the four PPases range from 0.83 to 1.14 Å. The mutual positions of the central β-barrel structure and α-helices are similar. However, the structure-based sequence alignment (Fig. 2 *A*) shows that Pho-PPase contains a single residue deletion in the loop formed by residues 25–30, two residues deleted in the loop formed by residues 111–114, and three residues deleted in the loop formed by residues 144–148.

Oligomeric structure

The Pho-PPase structure consists of three molecules in the asymmetric unit. The first subunit contains residues 4–171 out of a total of 178 residues; the second subunit contains residues 4–170; and the third subunit contains residues 4–170 with a Y170A substitution. Subunits 2 and 3 can be superimposed onto the first subunit with RMSDs of 0.61 Å and 0.65 Å, respectively. The three Pho-PPase monomers are related by a noncrystallographic threefold axis to form a tight trimer with an extensive subunit interface. As can been seen in Fig. 1 *B*, the main contact region concerns

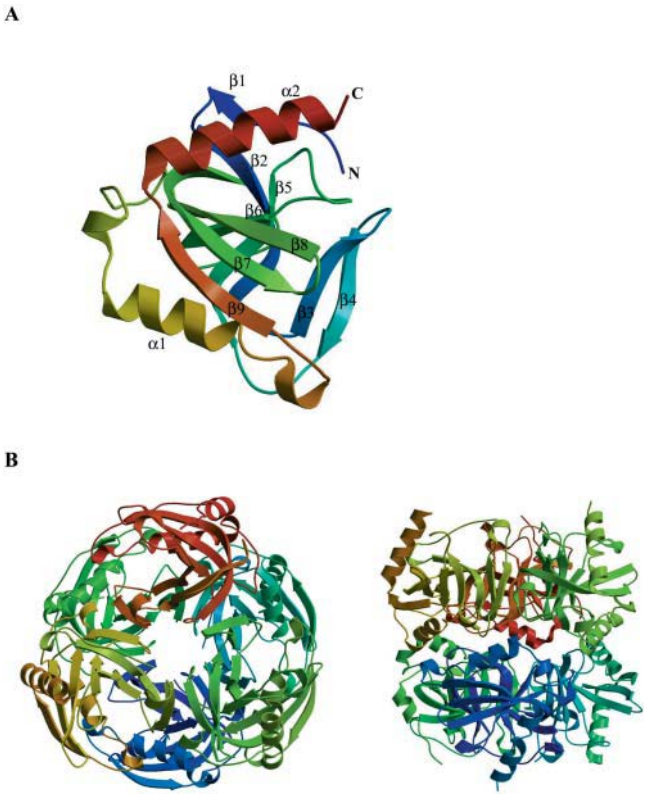

FIGURE 1 (*A*) A ribbon representation of the Pho-PPase monomer structure. The structure is colored from blue at the N-terminal to red at the C-terminus. Secondary structure elements have been labeled. (*B*) Top and side views of the Pho-PPase hexamer. Three subunits are related by a noncrystallographic threefold axis to form a trimer. Two trimers are related by a crystallographic twofold axis to form a tightly packed hexamer.

strands $\beta2$ and $\beta3$ of the β-barrel of one subunit and a β-hairpin (residues 78–85) of another subunit. There are many hydrophobic and hydrophilic interactions between subunits to enhance the enzyme stability, which will be discussed in further detail below.

From the crystallographic symmetry, the trimers are packed such as to form a tightly packed hexamer with twofold crystallographic symmetry. The intertrimer interactions in Pho-PPase are listed in Table 2. There are no large-scale distortions in monomeric structure, and the increased thermostability of Pho-PPase likely results from an increase in both hydrophilic and hydrophobic interactions. Intertrimer interactions are concentrated in strand $\beta3$, helix $\alpha1$, and the loop between $\beta8$ and $\alpha1$. The overall surface area buried in the hexamer is 2342 $Å^2$, which is comparable to the 2430 $Å^2$ buried by the T-PPase hexamer and higher than the 2090 $Å^2$ buried by the E-PPase hexamer (Salminen et al., 1996).

Active center

The active site cavity of Pho-PPase is formed between the β-barrel and helix $\alpha1$. The overall shape and size of the active site is very similar to those of *E. coli* and yeast PPases (Avaeva et al., 1997). It has been previously suggested that 17 residues might be involved in Mg^{2+} and PPi binding, 15 of which appear to be conserved in all sequences of soluble PPases known to date. Accordingly, all 15 of these conserved residues (E23, K31, E33, R44, Y52, Y56, D66, D68, D71, D98, D103, K105, Y140, K141, K146) also reside in the active-site cavity of Pho-PPase (Fig. 2). This implies that Pho-PPase will share the general catalytic mechanism of the PPase family in vivo, although the maximal enzyme activity of Pho-PPase is gained under a strong alkaline environment in vitro. Although Pho-PPase includes the conserved residues involved in Mg^{2+} binding, no Mg ion was observed in the Pho-PPase structure despite the addition of $MgCl_2$ during crystallization. Many crystals soaked in $MgCl_2$ were found to crack, which suggests that conformational changes may occur to prevent Mg^{2+} binding. A similar observation was made in the structure of E-PPase (Kankare et al., 1994).

The crystal structure of Pho-PPase reveals that K102, equivalent to residue E101 in E-PPase, is important for the active-site cavity and is located in the highly conserved region including the essential catalytic residues (numbered E97, D102, and K104 in E-PPase). Mutagenesis studies of E-PPase have shown that the enzyme activity of mutants E99D, D102E, and K104E almost disappeared or decreased rapidly (Hyytia et al., 2001). However, when E101 is replaced by a more negative aspartic acid residue, the enzyme activity of the mutant was increased by 10%. Alignment with other PPases shows that the residue located in the site is conserved as either an acidic or neutral residue. We propose that K102 is related to the alkaline optimal pH (10.3) of Pho-PPase, since only in such an alkaline environment would K102 not be positive. Further site-directed mutagenesis studies are in progress to investigate the role of K102.

Structural basis of thermostability in Pho-PPase

The free energy of stabilization of globular proteins is rather small. It lies in the range from 30 to 65 kJ/mol (Pfeil et al., 1986), which is equivalent to the energy contributed by a few hydrogen bonds, ion pairs, or hydrophobic interactions. The increase in free energy of stabilization observed for thermophilic proteins is of the same order of magnitude (Harris et al., 1980; Nojima et al., 1978). Recent structural studies also have identified several factors which are more often observed among thermophilic proteins and may account for their stability. These include an increased number of salt bridges or hydrogen bonds; optimized stability of helices, loops, and N- and C-termini; decreased solvent-exposed surface area; stronger interactions between the subunits in oligomers; and even an increased number of buried solvent molecules in hydrophilic cavities. We compare the present 2.66 Å resolution structure with three

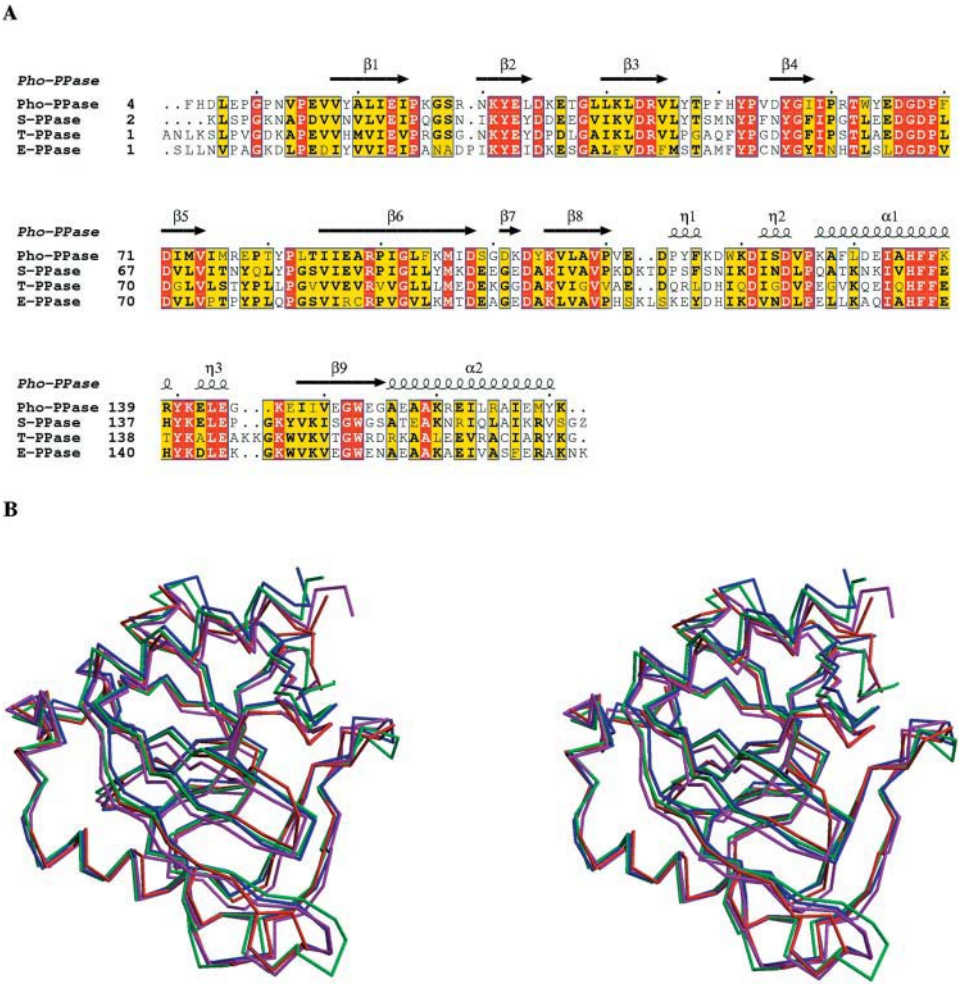

FIGURE 2 (*A*) A structure-based sequence alignment between Pho-PPase and three related PPase structures from *T. thermophilus* (T-PPase; PDB ID 2PRD) (Teplyakov et al., 1994), *S. acidocaldarius* (S-PPase; PDB ID 1QEZ) (Leppanen et al., 1999) and *E. coli* (E-PPase; PDB ID 1JFD) (Avaeva et al., 1997). Secondary structure elements are shown for Pho-PPase. (*B*) A stereodiagram showing the superposition of the four PPase structures. Structures are represented as C_α backbone traces. The coloring is as follows: red, Pho-PPase; blue, S-PPase; green, T-PPase; magenta, E-PPase. RMSD between C_α atoms of the four PPases range from 0.83 to 1.14 Å.

other PPases, to identify which factors may be important in stabilizing Pho-PPase.

Amino acid composition and thermostability

The amino acid composition of a protein has long been thought to be correlated with its thermostability. Compared with related PPase structures, the unusual amino acid composition of Pho-PPase would account for its extreme thermostability. The structure-based sequence alignment (Fig. 2) shows the amino acid sequences of the four PPases compared in this study. More charged residues were found in Pho-PPase, which is consistent with the idea that the number of ion pairs is an important determinant of protein thermostability. For instance, Pho-PPase contains a total number of eight arginine residues, which have a tendency to form multiple ion-pairs and H-bonds. Pho-PPase also contains 14 proline residues, which is the largest proportion in the four PPases studied here. Since proline residues affect local mobility of the chain by decreasing the conformational entropy of the unfolded state, the increased rigidity of the structure would be expected to increase the overall thermostability. Such stabilization by the introduction of proline residues into loop regions is a well-documented phenomenon (Matthews et al., 1987). A recent study of inorganic pyrophosphatase from thermophilic bacterium PS-3 was carried out in which proline residues were systematically replaced by alanines (Masuda et al., 2002). The authors found that most of the proline residues in PS-3 PPase play very important roles, and many of them are critical for the structural integrity of the protein. They also concluded that the thermostability of PS-3 PPase is profoundly related with its subunit structure.

The total number of hydrophobic residues (Gly, Ala, Val, Leu, Ile, Met, Phe, Trp, and Pro) is also highest in Pho-PPase. Compositional differences between the PPases are more pronounced among exposed sites. Pho-PPase has an increased number of exposed hydrophobic residues, which are presumably involved in oligomer formation and stability.

Interestingly, previous site-directed mutagenesis of E-PPase has shown that aromatic residues play a very important role for the thermostability (Hyytia et al., 2001). There are an increased number of aromatic residues found in Pho-PPase compared to three other PPases. In particular,

TABLE 2 Oligomeric interactions in E-PPase, T-PPase, S-PPase, and Pho-PPase

E-PPase a distance			T-PPase a distance			S-PPase b distance			Pho-PPase distance		
From	To	(Å)	From	To	(Å)	From	To	(Å)	From	To	(Å)
Intratrimeric A-B hydrophilic contacts											
S1	S36	2.79	N27	Y77	3.59	N16	K40	3.41	R28	E78	2.87
Y30	Q80	3.32	I39	V84	2.93	Y30	Y80	3.48	N29	E29	2.98
L39	V84	2.86	L41	V84	2.83	I39	V84	3.09	N29	E78	3.44
V41	V84	2.85	V44	E111	3.23	V41	V84	2.90	N29	P79	3.21
F44	L113	2.96	L45	R116	2.86	D42	T113	3.52	L40	I85	2.76
			L45	R116	2.93	V44	T113	2.78	K41	E87	3.39
			Q115	L144	3.56	Y46	K112	2.49	L42	I85	2.86
									R44	E112	2.82
									V45	G112	2.67
Intertrimeric A-D hydrophilic contacts											
S46	Q133	2.89	A48	F50	3.54	S48	N50	3.13	R28	E132	3.00
A48	F50	3.01	Q132	K148	3.00	S48 Oγ	N50	2.57	T48	H135	3.58
H136	D143	2.76	H136	T140	3.21	H136	H140	3.45	P49	H51	2.64
						H136	E143	2.66	H135	E142	2.48
						R139	E142	3.01	R138	E142	3.16
Intertrimeric A-E hydrophilic contacts											
N24	Y77	3.35	R116	E133	3.33				R77	E132	2.74
			R116	E133	3.27						
			R116	E133	2.82						
			Q115	Q132	3.33						
Intratrimeric A-B hydrophobic contacts											
L2	A28	3.92	P6	L36	3.91	L5	E36	4.03	F4	L39	4.07
D26	Y77	3.58	Y30	V83	3.98	L5	V38	4.05	R28	E78	3.84
P27	Y77	2.89	I39	V83	3.77	I28	Y77	3.78	N29	E78	4.10
I28	Y77	3.76	L41	V109	3.81	I28	L79	4.10	L39	L20	3.85
Y30	S83	3.99	R43	E111	3.98	Y30	Y80	3.51	L40	T84	3.91
L39	S83	3.94	V44	Y77	3.60	K40	V84	4.02	K41	E87	3.79
F40	V84	3.67	V44	V109	3.99	V41	L79	3.95	L42	M76	3.39
V41	L79	3.80	V44	D114	3.81	V41	T113	4.01	L42	I86	3.90
V41	L113	3.73	L80	P81	3.49	V44	Y77	3.72	K44	E112	3.78
F44	T75	3.40	Y46	P115	3.40				V45	M76	4.02
F44	P76	3.69	Y80	P81	3.63				V45	D113	3.53
F44	S114	3.99							Y47	P114	3.89
									P82	Y81	3.60
Intertrimeric A-D hydrophobic contacts											
P27	F50	3.99	A48	F50	3.79	T47	K133	3.38	R28	E132	3.87
S46	L129	3.86	A48	P52	4.00	S48	P52	4.09	T48	H135	3.62
T47	H136	3.57	A48	E133	3.80	S48	K133	4.02	T48	F50	3.24
A48	F50	3.73	Q49	H136	3.91	M49	M49	3.38	P49	F50	3.83
M49	M49	3.88	F50	F50	3.56	H136	L144	3.99	P49	E132	3.38
H136	L144	3.96	Q132	L144	3.98	H140	H140	3.25	P49	P53	4.05
H136	D143	3.37	T140	T140	3.76				F50	F50	4.06
H140	H140	3.61							H51	H51	3.47
									H135	L143	3.94
									R139	R139	3.51
Intertrimeric A-E hydrophobic contacts											
R116	E133	4.00	N76	N76	3.81–4.00				P114	A128	3.58
H119	P127	3.41	P115	A129	3.88–4.18				P114	A128	3.58
									Y115	P126	3.69
									Y115	A128	3.82

TABLE 3 Ionic interactions within a monomer

E-PPase distance			T-PPase distance			S-PPase distance			Pho-PPase distance		
From	To	(Å)	From	To	(Å)	From	To	(Å)	From	To	(Å)
E13	K162	3.66	K10	E86	3.69	D13	K162	2.64	E15	K161	3.16
D14	R86	2.79	E13	R166	4.12	K29	E31	2.81–3.93	K30	D43	3.12–3.76
D14	H110	3.66	K29	E31	3.86	D33	K40	2.57–5.60	K41	E87	3.96
K29	D42	2.7	D42	R43	2.11	E36	K40	3.20–5.99	R44	E144	3.12
R43	E154	2.71	R43	E145	4.02	R43	E145	2.43–3.20	R44	K112	2.82
D70	K104	2.87	E64	R171	2.89	R43	D42	2.85	D71	K105	3.72
K94	E101	3.47	D70	K104	3.03	E63	R164	2.83	R77	D113	3.26–3.72
D97	K142	3.55	E98	K99	4.5	E63	R171	3.90	K95	E154	4.00
D102	K104	3.07	E98	K142	2.58	E64	R171	3.88	K95	E155	3.16
H140	E139	3.43	D102	K104	3.02	D70	K104	2.94	D98	K141	3.98
H140	D143	2.93	D114	R116	3.13	E86	K110	3.22	D103	K105	3.76
E159	K162	3.76	D122	R158	3.84	R88	D111	3.11	K117	D118	3.15
			E98	K142	3.15	E145	K148	3.65	D118	K120	3.54
			E98	K148	3.50	E163	R166	4.04	K127	D131	3.16
					−4.09	D102	K104	2.91	E132	H135	3.73
						H140	E143	3.04	R139	E142	3.26
						E145	K148	2.46–4.06	R139	K138	3.46
									E144	K146	3.56
									E157	K160	3.24
									E157	K160	3.58
									E162	R165	2.22
									E169	K171	3.17

there is an increased frequency of phenylalanine and tyrosine residues in Pho-PPase, which are liable to form hydrophobic and aromatic interactions. Many of these aromatic residues are observed to form a cluster located at the bottom of the active site, and it is possible that stacking interactions involving the aromatic residues may contribute to enhanced thermostability of Pho-PPase.

The frequency of Asn (2) and Gln (0), which can be classed as thermolabile due to their tendency to undergo deamidation at high temperatures and therefore may be naturally discriminated against in thermostable proteins, is substantially reduced in Pho-PPase. Cysteine was also completely absent in Pho-PPase, which is easy to interpret since cysteine is highly sensitive to oxidation at high temperature. The frequency of glycine was not decreased, but changed in location. Interestingly, residue L83 of Pho-PPase is strictly conserved as a glycine in other PPases and is located in the short loop connecting β-strands 5 and 6, thus increasing the rigidity of the Pho-PPase structure.

Ionic interactions and thermostability

The importance of ion-pairs as determinants of protein thermostability was first highlighted by Perutz and Raidt (Perutz and Raidt, 1975) while comparing ferredoxin and hemoglobin structures, and ion-pairs were subsequently proposed to be important for the stability of a number of other thermostable proteins (Korndorfer et al., 1995; Walker et al., 1980). Indeed, several reports of high resolution structures of hyperthermophilic proteins show the number of ion-pairs in most of the hyperthermophilic proteins is higher than in their mesophilic counterparts (DeDecker et al., 1996; Hennig et al., 1995; Yip et al., 1995). A total of 12 ionic interactions are formed per monomer in the E-PPase short *c*-axis crystal form, which is equivalent to the number in T-PPase and lower than S-PPase (17 ionic pairs). In contrast, there are 28 ionic interactions scattered throughout the Pho-PPase monomer (Table 3). It appears that Pho-PPase is more stabilized by ion-pairs than T-PPase, E-PPase, and S-PPase. Although the number of ion-pairs is increased, the multicenter ion interactions are decreased. Interestingly, two long ion-pair networks are observed in Pho-PPase. The first ion-pair network is located in the C-terminus α2 helix, which may stabilize the C-terminus against thermal denaturation. The second ion network is located in the active center and involves residues 112–120 and 127–139. There is also an increase in the number of intrasubunit ion-pairs in Pho-PPase, and their involvement in complex networks mirrors that observed in the structure of the hyperthermophilic *Pyrococcus furiosus* glutamate dehydrogenase (Yip et al., 1995). The presence of ion-pair networks has also been observed in *Sulfolobus solfataricus* indole-3-glycerol phosphate synthase (Hennig et al., 1995) and the TATA-box binding protein from *P. furiosus* (DeDecker et al., 1996). The presence of ion-pair networks may be energetically favorable due to the shared entropic cost upon ion-pair formation (Nakamura, 1996).

The positioning of ion-pairs is also crucial (Daggett and Levitt, 1993). Loop or random coil regions of a protein tend to be the most flexible areas and are the most likely to deform at high temperatures. Compared with the three other PPases used in this study, Pho-PPase has three shorter loops, of

which two lie in the active site and the third sits at the molecular surface. In contrast, the loops found in E-PPase are longest of the four PPase structures. This observation, that the higher the growth temperature of the organism the shorter the protein loops, is consistent with high temperature molecular dynamic simulations on bovine pancreatic trypsin inhibitor, which revealed that loop and turn regions are likely to be the parts of the structure that unfold first during thermal denaturation. The structure of thermostable endocellulase (Sakon et al., 1996) also showed a reduced size of loop regions, and a similar strengthening of loop regions was also observed in *S. solfataricus* indole-3-glycerol phosphate synthase (Hennig et al., 1995).

Oligomerization and thermostability

In the PPase family, the oligomerization state stabilizes the conformation of the enzyme that binds substrate and vice versa (Baykov et al., 1995). The oligomeric packing appears to provide a general strategy for enhancing the thermostability of the PPase family. Pho-PPase seems to be a more tightly packed hexamer. The total number of oligomeric hydrophilic contacts is listed in Table 2. There are notably more intermonomer hydrophilic interactions in Pho-PPase compared with the three other PPases, resulting in tighter twofold A-E and threefold A-B interfaces. This effect can be measured by the accessible surface area (ASA) buried per monomer on oligomerization, which is 13.6% higher than for S-PPase and 12.1% higher than for E-PPase. These increased hydrophilic interactions may provide the extra energy necessary for stabilization. Similar observations are found in several thermophilic protein structures. For example, thermostability was attributed to improved subunit interfaces in L-lactate dehydrogenase from *Bacillus stearothermophilus* (Kallwass et al., 1992), malate dehydrogenase from *Thermus flavus* (Kelly et al., 1993), and ornithine carbamoyltransferase from *P. furiosus* (Villeret et al., 1998). Other studies also indicate that multimer formation and subunit interactions are critical for thermal stability of, for example, hemocyanin from the ancient tarantula *Eurypelma californicum* (Sterner et al., 1995), phosphoribosyl anthranilate isomerase from the hyperthermophile *Thermotoga maritima* (Hennig et al., 1997), GluDH from the hyperthermophile *P. furiosus* (Vetriani et al., 1998), and chorismate mutase from the thermophilic archaeon *Methanococcus jannaschii* (MacBeath et al., 1998).

Helices and thermostability

It has been previously reported that the helical conformation is stabilized by oppositely charged ion-pair interactions (i.e., Glu-Lys, Glu-Arg, Asp-Lys, Asp-Arg) in the positions (i,i+4) or (i,i+3) (Scholtz et al., 1993). The α1 helix contains a single ion-pair between the nonconserved residues K127-D131. Examination of the C-terminus α2 helix sequence indicates that Pho-PPase contains a significant increase in the number of charged residues compared with the other PPases (Fig. 2 *A*). These residues—E157, R161, E162, R165, E168, K171—are ideally distributed in the sequence to allow the formation of intrahelix ion-pairs. The structure of Pho-PPase shows three intrahelical ion-pairs formed between residues E157-R161, E162-R165, and E169-K172 (Table 3), which are not conserved in other PPase structures. These additional ion-pairs may be responsible for the increased thermostability of Pho-PPase by stabilizing the C-terminus and increasing its resistance to denaturation.

CONCLUSIONS

Comparison of the structure of Pho-PPase with T-PPase, S-PPase, and E-PPase has identified a number of determining factors for the thermostability of PPase enzymes. Pho-PPase is the most thermostable of the four structures, and this is reflected by the following factors. First, shorter and more convergent loops in the active site imply that the two loops are important for enzyme-substrate or ion binding under high temperature, and this is consistent with the idea that shorter loops are more helpful for thermostable PPases to stabilize the conformation of the enzyme-substrate complex. Second, the basis of thermostability in PPases is generally believed to be related to the oligomer structure; Pho-PPase has an increase in the number of both intermonomer and intertrimer interactions, and this results in tighter packing of the hexamer. Third, an increase in the number of ion-pairs—in particular, two ion-pair networks—helps to stabilize the structure. One ion-pair network in the loop connecting strand β8 and helix α1 may help to stabilize the conformation of the active site, while another network in helix α2 may stabilize the C-terminus and increase its resistance to thermal denaturation. The structure of Pho-PPase and the comparison between PPase structures should stimulate further kinetic and structural studies of enzymes adapted to extreme temperature and prove helpful for future biotechnology applications.

We are grateful to Feng Gao for help with data collection.

This research was supported by the following grants: Project ''973'' G1999075602; Ministry of Science and Technology (MOST) 2002BA711A12.

REFERENCES

Avaeva, S., S. Kurilova, T. Nazarova, E. Rodina, N. Vorobyeva, V. Sklyankina, O. Grigorjeva, E. Harutyunyan, V. Oganessyan, K. Wilson, M. Dauter, R. Huber, and T. Mather. 1997. Crystal structure of Escherichia coli inorganic pyrophosphatase complexed with SO4(2-). Ligand-induced molecular asymmetry. *FEBS Lett.* 410:502–508.

Baykov, A. A., V. Y. Dudarenkov, J. Kapyla, T. Salminen, T. Hyytia, V. N. Kasho, S. Husgafvel, B. S. Cooperman, A. Goldman, and R. Lahti. 1995. Dissociation of hexameric Escherichia coli inorganic pyrophosphatase into trimers on His-136→Gln or His-140→Gln substitution and its effect on enzyme catalytic properties. *J Biol Chem.* 270:30804–30812.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D.* 54:905–921.

Chen, J., A. Brevet, M. Fromant, F. Leveque, J. M. Schmitter, S. Blanquet, and P. Plateau. 1990. Pyrophosphatase is essential for growth of Escherichia coli. *J. Bacteriol.* 172:5686–5689.

Daggett, V., and M. Levitt. 1993. Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* 232:600–619.

DeDecker, B. S., R. O'Brien, P. J. Fleming, J. H. Geiger, S. P. Jackson, and P. B. Sigler. 1996. The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J. Mol. Biol.* 264:1072–1084.

Hansen, T., C. Urbanke, V. M. Leppanen, A. Goldman, K. Brandenburg, and G. Schafer. 1999. The extreme thermostable pyrophosphatase from Sulfolobus acidocaldarius: enzymatic and comparative biophysical characterization. *Arch. Biochem. Biophys.* 363:135–147.

Harris, J. I., J. D. Hocking, M. J. Runswick, K. Suzuki, and J. E. Walker. 1980. D-glyceraldehyde-3-phosphate dehydrogenase. The purification and characterisation of the enzyme from the thermophiles Bacillus stearothermophilus and Thermus aquaticus. *Eur. J. Biochem.* 108:535–547.

Heikinheimo, P., V. Tuominen, A. K. Ahonen, A. Teplyakov, B. S. Cooperman, A. A. Baykov, R. Lahti, and A. Goldman. 2001. Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase. *Proc. Natl. Acad. Sci. USA.* 98:3121–3126.

Hennig, M., B. Darimont, R. Sterner, K. Kirschner, and J. N. Jansonius. 1995. 2.0 A structure of indole-3-glycerol phosphate synthase from the hyperthermophile Sulfolobus solfataricus: possible determinants of protein stability. *Structure.* 3:1295–1306.

Hennig, M., R. Sterner, K. Kirschner, and J. N. Jansonius. 1997. Crystal structure at 2.0 A resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile Thermotoga maritima: possible determinants of protein stability. *Biochemistry.* 36:6009–6016.

Hyytia, T., P. Halonen, A. Salminen, A. Goldman, R. Lahti, and B. S. Cooperman. 2001. Ligand binding sites in Escherichia coli inorganic pyrophosphatase: effects of active site mutations. *Biochemistry.* 40:4645–4653.

Ichiba, T., T. Shibasaki, E. Iizuka, A. Hachimori, and T. Samejima. 1998. Cation-induced thermostability of yeast and Escherichia coli pyrophosphatases. *Biochem. Cell Biol.* 66:25–31.

Jones, T. A., J. Y. Zou, S. W. Cowan, and M. Kjeldgaard. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A.* 47:110–119.

Kallwass, H. K., W. K. Surewicz, W. Parris, E. L. Macfarlane, M. A. Luyten, C. M. Kay, M. Gold, and J. B. Jones. 1992. Single amino acid substitutions can further increase the stability of a thermophilic L-lactate dehydrogenase. *Protein Eng.* 5:769–774.

Kankare, J., G. S. Neal, T. Salminen, T. Glumhoff, B. S. Cooperman, R. Lahti, and A. Goldman. 1994. The structure of E. coli soluble inorganic pyrophosphatase at 2.7 A resolution. *Protein Eng.* 7:823–830.

Kelly, C. A., M. Nishiyama, Y. Ohnishi, T. Beppu, and J. J. Birktoft. 1993. Determinants of protein thermostability observed in the 1.9-A crystal structure of malate dehydrogenase from the thermophilic bacterium Thermus flavus. *Biochemistry.* 32:3913–3922.

Korndorfer, I., B. Steipe, R. Huber, A. Tomschy, and R. Jaenicke. 1995. The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the thermophilic bacterium Thermotoga aquaticus DNA polymerase at 2.5 A resolution. *J. Mol. Biol.* 246:511–521.

Lahti, R. 1983. Microbial inorganic pyrophosphatases. *Microbiol. Rev.* 47:169–179.

Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283–291.

Leppanen, V. M., H. Nummelin, T. Hansen, and R. Lahti. 1999. Sulfolobus acidocaldarius inorganic pyrophosphatase: structure, thermostability, and effect of metal ion in an archael pyrophosphatase. *Protein Sci.* 8:1218–1231.

MacBeath, G., P. Kast, and D. Hilvert. 1998. A small, thermostable, and monofunctional chorismate mutase from the archaeon Methanococcus jannaschii. *Biochemistry.* 37:10062–10073.

Masuda, H., T. Uchiumi, M. Wada, T. Ichiba, and A. Hachimori. 2002. Effects of replacement of prolines with alanines on the catalytic activity and thermostability of inorganic pyrophosphatase from thermophilic bacterium PS-3. *J. Biochem. (Tokyo).* 131:53–58.

Matthews, B. W., H. Nicholson, and W. J. Becktel. 1987. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. USA.* 84:6663–6667.

Nakamura, H. 1996. Roles of electrostatic interaction in proteins. *Q. Rev. Biophys.* 29:1–90.

Nojima, H., K. Hon-Nami, T. Oshima, and H. Noda. 1978. Reversible thermal unfolding of thermostable cytochrome c-552. *J. Mol. Biol.* 122:33–42.

Otwinowski, Z., and W. Minor. 1997. Processing of x-ray diffraction data collected in oscillation mode. *In*: Macromolecular Crystallography, part A. C. W. Carter Jr. and R. M. Sweet, editors. Academic Press. 307–326.

Perutz, M., and H. Raidt. 1975. Stereochemical basis of heat thermostability in bacterial ferredoxins and in hemoglobin A2. *Nature.* 255:256–259.

Pfeil, W., V. E. Bychkova, and O. B. Ptitsyn. 1986. Physical nature of the phase transition in globular proteins. Calorimetric study of human alpha-lactalbumin. *FEBS Lett.* 198:287–291.

Richter, O. M., and G. Schafer. 1992. Purification and enzymic characterization of the cytoplasmic pyrophosphatase from the thermoacidophilic archaebacterium *Thermoplasma acidophilum*. *Eur. J. Biochem.* 209:343–349.

Sakon, J., W. S. Adney, M. E. Himmel, S. R. Thomas, and P. A. Karplus. 1996. Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. *Biochemistry.* 35:10648–10660.

Salminen, T., A. Teplyakov, J. Kankare, B. S. Cooperman, R. Lahti, and A. Goldman. 1996. An unusual route to thermostability disclosed by the comparison of *Thermus thermophilus* and *Escherichia coli* inorganic pyrophosphatases. *Protein Sci.* 5:1014–1025.

Scholtz, J. M., H. Qian, V. H. Robbins, and R. L. Baldwin. 1993. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry.* 32:9668–9676.

Sterner, R., T. Vogl, H. J. Hinz, F. Penz, R. Hoff, R. Foll, and H. Decker. 1995. Extreme thermostability of tarantula hemocyanin. *FEBS Lett.* 364:9–12.

Teplyakov, A., G. Obmolova, K. S. Wilson, K. Ishii, H. Kaji, T. Samejima, and I. Kuranova. 1994. Crystal structure of inorganic pyrophosphatase from *Thermus thermophilus*. *Protein Sci.* 3:1098–1107.

Vander Horn, P. B., M. C. Davis, J. J. Cunniff, C. Ruan, B. F. McArdle, S. B. Samols, J. Szasz, G. Hu, K. M. Hujer, S. T. Domke, S. R. Brummet, R. B. Moffet, and C. W. Fuller. 1997. Thermo Sequenase DNA polymerase and *T. acidophilum* pyrophosphatase: new thermostable enzymes for DNA sequencing. *Biotechniques* 22(4):758–762, 764–765.

Vetriani, C., D. L. Maeder, N. Tolliday, K. S. Yip, T. J. Stillman, K. L. Britton, D. W. Rice, H. H. Klump, and F. T. Robb. 1998. Protein thermostability above 100 degreesC: a key role for ionic interactions. *Proc. Natl. Acad. Sci. USA.* 95:12300–12305.

Villeret, V., B. Clantin, C. Tricot, C. Legrain, M. Roovers, V. Stalon, N. Glansdorff, and J. Van Beeumen. 1998. The crystal structure of *Pyrococcus furiosus* ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures. *Proc. Natl. Acad. Sci. USA.* 95:2801–2806.

Walker, J. E., A. J. Wonacott, and J. I. Harris. 1980. Heat stability of a tetrameric enzyme, D-glyceraldehyde-3-phosphate dehydrogenase. *Eur. J. Biochem.* 108:581–586.

Yip, K. S., T. J. Stillman, K. L. Britton, P. J. Artymiuk, P. J. Baker, S. E. Sedelnikova, P. C. Engel, A. Pasquo, R. Chiaraluce, and V. Consalvi. 1995. The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure.* 3:1147–1158.