Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A monomeric TIM-barrel structure from *Pyrococcus furiosus* is optimized for extreme temperatures

Heidi Repo,[a] Jesper S. Oeemig,[a] Janica Djupsjöbacka,[a] Hideo Iwaï[a]* and Pirkko Heikinheimo[b]*

[a]Institute of Biotechnology, University of Helsinki, PO Box 65, FI-00014 Helsinki, Finland, and [b]Department of Biochemistry and Food Chemistry, University of Turku, FI-20014 Turku, Finland

Correspondence e-mail: hideo.iwai@helsinki.fi, pirkko.heikinheimo@utu.fi

Received 21 February 2012
Accepted 28 August 2012

PDB Reference: *Pf*TrpF, 4aaj

The structure of phosphoribosyl anthranilate isomerase (TrpF) from the hyperthermophilic archaeon *Pyrococcus furiosus* (*Pf*TrpF) has been determined at 1.75 Å resolution. The *Pf*TrpF structure has a monomeric TIM-barrel fold which differs from the dimeric structures of two other known thermophilic TrpF proteins. A comparison of the *Pf*TrpF structure with the two known bacterial thermophilic TrpF structures and the structure of a related mesophilic protein from *Escherichia coli* (*Ec*TrpF) is presented. The thermophilic TrpF structures contain a higher proportion of ion pairs and charged residues compared with the mesophilic *Ec*TrpF. These residues contribute to the closure of the central barrel and the stabilization of the barrel and the surrounding α-helices. In the monomeric *Pf*TrpF conserved structural water molecules are mostly absent; instead, the structural waters are replaced by direct side-chain–main-chain interactions. As a consequence of these combined mechanisms, the *P. furiosus* enzyme is a thermodynamically stable and entropically optimized monomeric TIM-barrel enzyme which defines a good framework for further protein engineering for industrial applications.

1. Introduction

Phosphoribosyl anthranilate isomerase (TrpF; EC 5.3.1.24) is a sugar isomerase involved in tryptophan biosynthesis in most bacteria (Henn-Sax *et al.*, 2002). It converts *N*-(5′-phosphoribosyl)-anthranilate to 1-(*o*-carboxyphenylamino)-1-deoxyribulose 5-phosphate (Sterner *et al.*, 1996). In *Escherichia coli*, a related dual-function phosphoribosyl anthranilate performs the two last steps of tryptophan synthesis; its C-terminal domain corresponds to the monofunctional TrpF (*Ec*TrpF; Wilmanns *et al.*, 1992).

Pyrococcus furiosus is a thermophilic archaeon isolated from marine solfataric fields (Vieille & Zeikus, 2001). Proteins with extreme stabilities are essential for biotechnological, biomedical and industrial applications (Vieille & Zeikus, 2001). The general strategy for increasing protein stability is to either stabilize the folded state by improving hydrophobic packing and increasing hydrogen bonds and ion pairs, or to destabilize the unfolded state using disulfide bonds or circular peptides (Razvi & Scholtz, 2006; Vieille & Zeikus, 2001). However, increasing protein stability without disturbing functionality is not a trivial task. Thus, proteins isolated from thermophiles are of great interest to protein engineers as they are natively thermostable and functional at elevated temperatures.

TrpF structures from two thermophilic bacteria are known: *Thermotoga maritima* (*Tm*TrpF; PDB entry 1lbm; Henn-Sax *et al.*, 2002) and *Thermus thermophilus* HB8 (*Tt*TrpF; PDB entry 1v5x; Taka *et al.*, 2005). Both of these enzymes function as and

Table 1
Data-collection and refinement statistics for *Pf*TrpF.

Values in parentheses are for the last resolution shell.

Wavelength (Å)	0.93950
Space group	$P6_1$
Unit-cell parameters (Å)	$a = b = 105.5$, $c = 34.5$
No. of molecules in asymmetric unit	1
Solvent content (%)	49
Data collection	
Resolution (Å)	20.0–1.75 (1.79–1.75)
$R_{\rm meas}$† (%)	4.4 (64.9)
Completeness (%)	99.9 (98.8)
No. of unique observations	22537
Average multiplicity	6.5 (5.3)
$I/\sigma(I)$‡	22.7 (2.9)
Refinement	
No. of reflections	21391
No. of reflections in test set	1126
Final R factors§	
$R_{\rm work}$	0.219
$R_{\rm free}$	0.258
Final structure	
No. of protein atoms	1561
No. of waters	41
No. of other atoms	15
Mean B value (Å^2)	42.5
R.m.s.d. from ideal¶	
Bond lengths (Å)	0.007
Bond angles (°)	1.1
Ramachandran plot	
Residues in favoured region (%)	98

† The redundancy-independent R factor on intensities ($R_{\rm meas}$) is defined in Diederichs & Karplus (1997). ‡ $I/\sigma(I)$ is the mean of $\langle$intensity/σ(intensity)$\rangle$ of unique reflections after merging the symmetry-related observations (Kabsch, 2010). § R factor = $\sum_{hkl} \big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure factors, respectively. $R_{\rm free}$ is calculated using a test set of reflections as described in Brünger (1992). ¶ Ideal values from Engh & Huber (1991)

were crystallized as dimers. The TrpF structure has a TIM-barrel fold, which is one of the most frequent enzyme folds (Wierenga, 2001; Höcker *et al.*, 2001). The TIM-barrel fold is also a versatile scaffold for the *de novo* design of enzymes (Höcker *et al.*, 2001; Röthlisberger *et al.*, 2008). A better understanding of extremely thermostable TIM barrels will provide a structural basis for enzyme engineering for biotechnological and industrial applications (Wierenga, 2001).

We have now determined the structure of TrpF from the hyperthermophilic archaeon *P. furiosus* (*Pf*TrpF; UniProt Q8U092) at 1.75 Å resolution. *Pf*TrpF is a monomer in solution and has an alternative mechanism for stability compared with dimeric TrpFs. We also present a comparison of known thermophilic TrpF structures with that of a related mesophilic protein from *E. coli*. We shed light on the structural basis for the intrinsic thermostability of the monomeric *Pf*TrpF, which is highly relevant for further engineering and design based on a thermostable monomeric TIM-barrel template structure.

2. Materials and methods

2.1. Protein expression and purification

The gene for *Pf*TrpF was amplified from genomic DNA of *P. furiosus* (ATCC 43587) using the oligonucleotides HK120, 5′-ATC ATA TGT TCG TAA AAA TAT GCG G, and HK121, 5′-CAG AAG CTT ACC AAA CCA CAT TTT TGG CCC. The PCR product was cloned into vector pHYRSF1-02 derived from pRSF-1b (Novagen) using *Nde*I and *Kpn*I restriction sites, resulting in the plasmid pJDJRSF02. The plasmid bearing the *trpF* gene encoding phosphoribosylanthranilate isomerase (*Pf*TrpF) was transfected into *E. coli* ER2566 cells (New England Biolabs) for protein expression. The cells were grown in 2 l LB medium and were induced with 0.5 m*M* IPTG at an OD_{600} of 0.7. The cells were harvested by centrifugation after 4 h induction, resuspended in 50 m*M* sodium phosphate buffer pH 8, 300 m*M* NaCl and heated at 363 K for 20 min. The cell lysate was cleared by centrifugation at 38 700*g* for 1 h. The clear supernatant was loaded onto a HisTrap column (GE Healthcare). After washing with 50 m*M* sodium phosphate, 10 m*M* imidazole pH 8, 300 m*M* NaCl, the protein was eluted using a linear gradient of 50–250 m*M* imidazole in the same buffer. Fractions containing *Pf*TrpF according to SDS–PAGE were pooled and dialysed overnight against phosphate-buffered saline (PBS). The solution was concentrated and loaded onto Superdex 75 (GE Healthcare) with PBS for further purification. Prior to crystallization, the protein was dialyzed against 20 m*M* MES pH 6.0 and concentrated to 10 mg ml^{-1} using a 10 kDa cutoff Microcon centrifugal filter device (Millipore).

An NMR sample was prepared from cells induced in M9 medium supplemented with $^{15}NH_4Cl$ as the sole nitrogen source. ^{15}N-labelled *Pf*TrpF was purified as described above, concentrated to 0.2 m*M* (5 mg ml^{-1}) in 10 m*M* sodium phosphate pH 6.0 using a centrifugal filter device (Millipore) and transferred to a Shigemi microtube for NMR measurements.

2.2. Crystallization and data collection

Crystallization conditions for *Pf*TrpF were screened using automation at the Crystallization Facility of the Institute of Biotechnology, University of Helsinki. Initial crystals were obtained from six of the 96 conditions in the Helsinki Random I screen (http://www.biocenter.helsinki.fi/bi/X-ray/automation/services.htm); these conditions were further optimized in order to produce data-collection-quality crystals. Crystals for data collection were obtained from hanging-drop vapour-diffusion experiments at room temperature in which 10 mg ml^{-1} *Pf*TrpF was mixed in a 1:1 ratio with 1.8 *M* ammonium sulfate in 0.1 *M* sodium acetate buffer pH 4.5. Rod-shaped crystals grew in 1 d. The cryoprotectant solution consisted of 25% glycerol in the crystallization solution.

*Pf*TrpF crystallized in space group $P6_1$, with unit-cell parameters $a = b = 105.5$, $c = 34.5$ Å (Table 1). Data were collected to 1.75 Å resolution at 100 K on the ID14-4 beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. High-resolution and low-resolution data sets were collected separately from a single crystal using a wavelength of 0.93950 Å. Both data sets contained images corresponding to 90° of reciprocal space and were processed using the *XDS* package (Kabsch, 2010) before scaling them together in *XSCALE* (Kabsch, 2010).

2.3. Size-exclusion chromatography

Size-exclusion gel chromatography was performed on an ÄKTApurifier 10 system with 0.25 mm tubing (Amersham Pharmacia Biotech) using a Superdex 200 5/150 column. The column was equilibrated at 0.3 ml min^{-1} with 50 m*M* Tris, 150 m*M* NaCl pH 7.4. A sample volume of 10 µl was injected. The concentration of *Pf*TrpF was 1.8 mg ml^{-1}. The elution volume of TrpF was compared with the following gel-filtration standards (Sigma): cytochrome *c* (12.4 kDa), carbonic anhydrase (29 kDa) and bovine serum albumin (66 kDa).

2.4. NMR relaxation measurements

NMR measurements were performed at 308 K on a Varian INOVA 600 or 800 MHz spectrometer equipped with a triple-resonance cryogenic probe head. $T_1(^{15}N)$ relaxation rates were determined with T_1 relaxation delays of 10, 30, 50, 100, 150, 200 and 250 ms. $T_2(^{15}N)$ relaxation rates were obtained based on a CPMG-type sequence with T_2 relaxation delays of 10, 30, 50, 70 and 90 ms. T_1 and T_2 relaxation rates were determined from the peak volumes using *CcpNmr Analysis* (Vranken *et al.*, 2005). T_1 and T_2 relaxation rates were determined for 40 residues and the T_1/T_2 ratio was used for the calculation of a global correlation time (τ_c) using an isotopic model and the *DASHA* software (Orekhov *et al.*, 1995). Only well separated peaks from NH groups in HSQC spectra were selected for relaxation analysis.

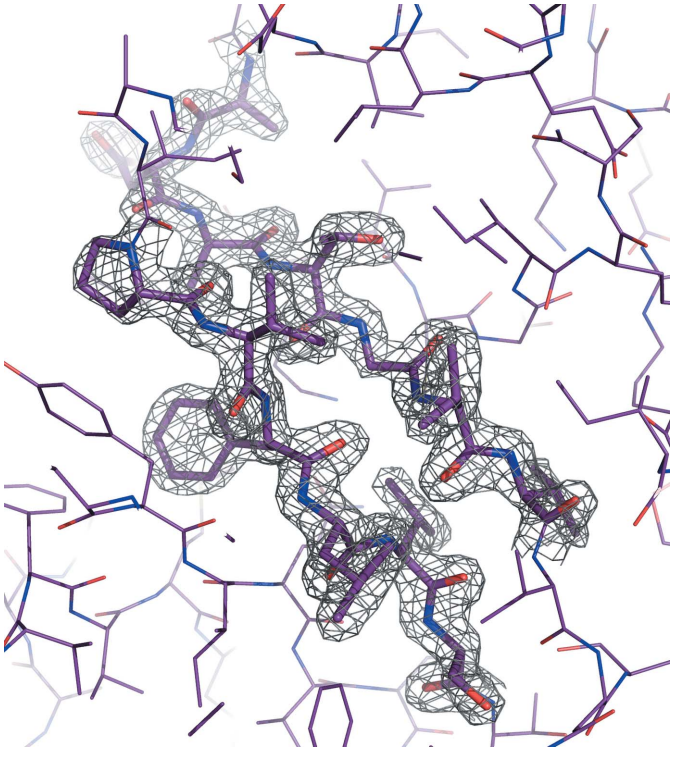

Figure 1
Fit of the final model to the density. The σ_A-weighted $2mF_o - DF_c$ electron density for residues Ala21–Val27 and Pro51–Ser56 is displayed at 0.47 e Å^{-3} (2σ).

Fast-HYDRONMR was used for estimation of correlation times based on the experimental crystal structures (Ortega & Garcia de la Torre, 2005).

2.5. Crystal structure solution and refinement

The X-ray crystallographic structure of *Pf*TrpF was determined by the molecular-replacement method with *MOLREP* in the *CCP*4 suite v.6.1.2 (Winn *et al.*, 2011) using the structure of *Tm*TrpF (PDB entry 1lbm; Henn-Sax *et al.*, 2002) as a search model. 5% of the observations were set aside for cross-validation analysis (Brünger, 1992). The molecular-replacement solution was further improved automatically with *ARP*/*wARP* in the *CCP*4 suite (Winn *et al.*, 2011). The model was manually completed using *Coot* (Emsley & Cowtan, 2004) and was refined with *REFMAC*5 (Murshudov *et al.*, 2011). Data-collection and refinement statistics are presented in Table 1. The free *R* factor (Brünger, 1992) converged to 0.258 and R_{work} to 0.219 in the final model (Table 1). The final fit to the electron-density map is shown in Fig. 1.

2.6. Structural analysis

The stereochemistry of the structure was analysed using *MolProbity* (Chen *et al.*, 2010). 98% of the residues were in the most favoured areas of the Ramachandran plot and none were in the disallowed region (Table 1). The TrpF structures were superimposed on each other using the *Secondary Structure Matching* (*SSM*) program as implemented in *Coot* (Emsley & Cowtan, 2004).

Salt bridges were analysed using the *Protein Structural Analysis Package* server (Balamurugan *et al.*, 2007). Potential ion pairs are defined as a distance of <4 Å between atoms of opposite charge from Arg, Lys, Glu and Asp residues. Owing to the weak contribution of histidine to the overall stability of a protein, these interactions were not included in the analysis of charged interactions. All of the interactions identified by the server were displayed in the *PyMOL* molecular-graphics system (v.1.2r2; http://www.pymol.org) and checked manually. We counted bidentate connections between arginine and acidic residues as a single salt bridge since the shared charge is only equal to 1. Solvation energy and inaccessible charged residues were analysed for each monomer using the *Protein Interactions, Surfaces and Assemblies* (*PISA*) server (Krissinel & Henrick, 2007; http://www.ebi.ac.uk/msd-srv/prot_int/). The accessible surface area (ASA), dimer and crystal interfaces were analysed by the *PISA* server (Krissinel & Henrick, 2007).

Cavity volumes in structures were analysed with *VOIDOO* (Kleywegt & Jones, 1994). Internal water molecules were identified with the program *DOWSER* (Zhang & Hermans, 1996). All structures were analysed as monomers.

3. Results

3.1. Overall structure

The monomer of *Pf*TrpF has a typical TIM-barrel fold with an eight-strand α/β barrel. The active-site and C-terminal end of the barrel binds two sulfate ions. The first one binds in a

similar position as the product analogue in the *Tm*TrpF structure (PDB entry 1lbm; Henn-Sax *et al.*, 2002). The other sulfate ion is also in the active-site region, with the central S atom only 2.3 Å away from the terminal O atom of the product-analogue molecule. The residues between Ala134 and Asp142 in the loop that follows β6 and folds over the active site are missing from the electron density and thus were not modelled. The N-terminal end of the barrel is well ordered. N-terminal residues are visible from the first purification-tag residue, but two last residues are missing from the C-terminus. The protein model is built of 200 protein residues, three sulfate ions and 41 water molecules, a total of 1617 atoms (Table 1).

3.2. *Pf*TrpF is a monomer in solution

Unlike the other reported TrpFs from thermophilic organisms, *Pf*TrpF behaves as a monomer in size-exclusion chromatography, NMR and crystal packing.

In size-exclusion chromatography, *Pf*TrpF elutes at an exclusion volume corresponding to a molecular weight of 19.5 kDa. Size-exclusion chromatography of *Pf*TrpF thus supports a monomer rather than a dimer in solution.

In addition to gel filtration, we used NMR relaxation analysis to characterize the oligomeric state of *Pf*TrpF in solution. The rotational correlation time caused by Brownian rotation for a spherical molecule of radius in a liquid of viscosity η can be defined as

$$\tau_c = \frac{4\pi a^3 \eta}{3kT} = \frac{V\eta}{kT}, \qquad (1)$$

where V is the volume of the molecule (Creighton, 1994). The volume of a typical monomeric protein is proportional to the molecular weight ($V = 1.27\mathrm{MW}$ Å^3 Da^{-1}; Creighton, 1994). As the nuclear relaxation time T_2 is inversely proportional to the rotational correlation time τ_c, the nuclear relaxation analysis provides an estimation of the molecular weight of a protein.

As a rule of thumb, for a given molecule the rotational correlation time in nanoseconds is approximately half of the molecular weight in kDa (Reddy & Rainey, 2010). In the case of *Pf*TrpF (MW 25 025.8 Da, including the His tag) the estimated correlation time would be around 11.5–12.5 ns for a monomeric form at room temperature, although even a small anisotropy in the molecular shape could affect the estimation of the relaxation analysis (Luginbühl *et al.*, 1997). From the *Pf*TrpF crystal structure, the rotational correlation time at 308 K calculated by the program *Fast-HYDRONMR* is 14.0 ns for a monomer or 27.1 ns for the related *Tf*TrpF dimer (PDB entry 1v5x; Taka *et al.*, 2005).

We used 40 well separated non-assigned peaks in the ^{15}N-HSQC spectrum of *Pf*TrpF recorded at a ^{1}H frequency of 800 MHz and excluded peaks outside the standard deviation. The rotational correlation time calculated by *DASHA* for *Pf*TrpF is 12.7 ± 2.7 ns. Using 32 peaks recorded with a 600 MHz NMR spectrometer, we obtained a rotational correlation time of 15.8 ± 4.9 ns. Thus, the correlation times from the ^{15}N relaxation-time analysis are in good agreement with the hydrodynamic calculation for a *Pf*TrpF monomer.

The crystal lattice of *Pf*TrpF is in good agreement with the monomeric nature of the enzyme. The lattice forms two

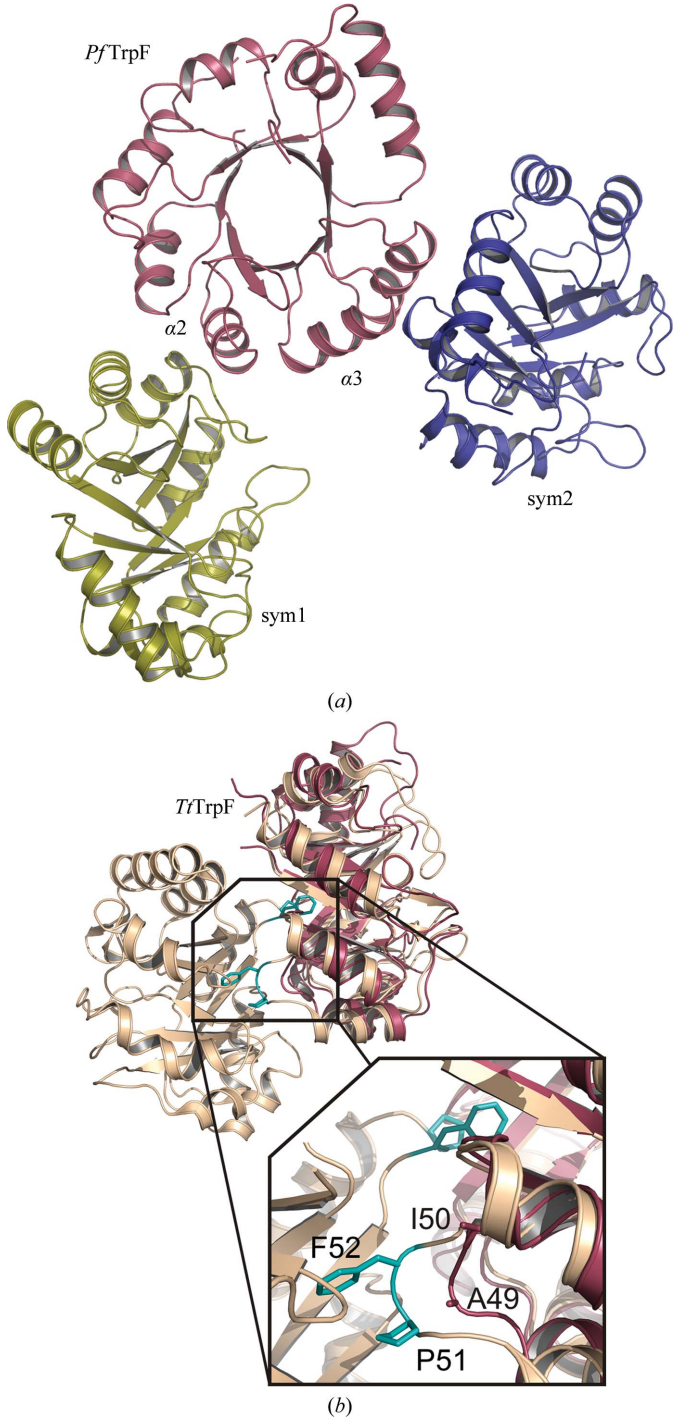

Figure 2
Symmetry contacts in *Pf*TrpF are smaller than the dimer interface of *Tf*TrpF. (*a*) A monomer of *Pf*TrpF is displayed in red and crystallographic symmetry mates are displayed in olive green (sym1) and blue (sym2). (*b*) A 90° rotation of (*a*) showing the *Tf*TrpF dimer superimposed on the *Pf*TrpF crystal structure. The enlarged view shows the dimer-interface loop Pro51-Phe52 (displayed in cyan) which is essential in holding the *Tf*TrpF dimer together. This structure is not conserved between *Pf*TrpF and *Tf*TrpF (the corresponding residues are Ala49 and Ile50 in *Pf*TrpF).

contact surfaces, both of which are small and weak. In the contact surface with the first symmetry mate, Glu39 and Glu46 from the α2 helix form a salt bridge to Arg35′ in a C-terminal barrel loop and Arg70 forms an interaction with a symmetry-related Glu15′ and, through a sulfate molecule, with Lys40′ (*Pf*TrpF in red and sym1 in olive green in Fig. 2*a*). The second symmetry-generated contact is mainly held together by hydrophobic packing of Phe61 from the α3 helix with Pro155′ at the N-terminal end of the barrel (*Pf*TrpF in red and sym2 in blue in Fig. 2*a*). The interfaces position the two barrels sideways at an angle relative to each other and in a different manner compared with the other TrpF crystal structures, which form native dimers. These dimers are stabilized by an interlocking loop which is not conserved in *Pf*TrpF (Fig. 2*b*). Only 5% of the total protein solvent-accessible area is buried in the larger symmetry-generated interface in the *Pf*TrpF crystal lattice, and one of the interfaces (but not the other) contains charged interactions. The area of the *Pf*TrpF (symmetry) interface is less than half of the largest (dimer) interface area in other thermophilic bacterial TrpF structures (Table 2).

Table 2
Structural comparison and statistics for the analysed TrpF structures.

All values are given per monomer. The solvation energy of folding is given as ΔG for 'all structure' from the *PISA* server (Krissinel & Henrick, 2007). This value reports the solvation-energy difference between the observed and reference states, which corresponds to the solvation energy of folding. *DOWSER* waters are theoretical buried structural waters which are predicted to make a stabilizing contribution to the protein structure. 1 cal = 4.186 J.

	*Pf*TrpF	*Tt*TrpF	*Tm*TrpF	*Ec*TrpF
Functional unit	Monomer	Dimer	Dimer	Domain
Sequence length†	203	203	205	198
Pro/Gly†	8/11	18/18	11/11	5/21
Arg†	11	20	15	8
Glu/Asp (total)†	19/8 (27)	21/5 (26)	16/12 (28)	10/12 (22)
Gln/Asn (total)†	3/10 (13)	4/1 (5)	2/8 (10)	13/9 (21)
MW (kDa)/volume (Å^3)†	22.3/27700	21.6/26600	21.9/26900	21.1/25500
pI†	5.95	6.35	5.93	5.84
Ion bonds (excluding His)†	13 (11)	13 (12)	15 (13)	10 (7)
Solvation energy of folding (kcal mol^{-1})	−189	−201	−200	−176
Largest interface (Å^2)	462	1254	1435	na
ASA (Å^2)	9180	9360	9480	9100
Inaccessible charged residues	0	1 (Lys101)	0	0
DOWSER waters	4	9	9	11

† These data were analysed using the *Protein Structure Analysis* package (http://iris.physics.iisc.ernet.in/psap/; Laskowski *et al.*, 2005).

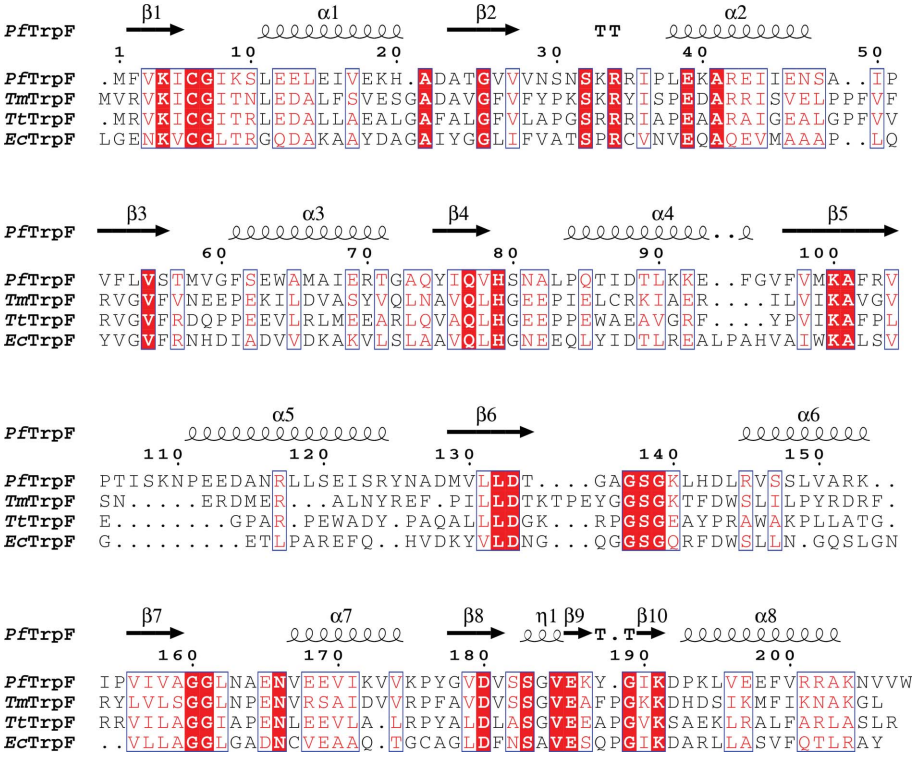

Figure 3
Structure-based sequence alignment of TrpF proteins. The initial alignment was made by *ClustalW* (Larkin *et al.*, 2007) and then manually edited to match the structural conservation. Secondary structures were analysed for the reference structure *Pf*TrpF in *ESPript* (Gouet *et al.*, 1999). The α5 helix is replaced by a loop in *Ec*TrpF. The aligned sequences are *Tt*TrpF (PDB entry 1v5x; Taka *et al.*, 2005), *Tm*TrpF (PDB entry 1lbm; Henn-Sax *et al.*, 2002) and *Ec*TrpF (PDB entry 1pii; Wilmanns *et al.*, 1992).

3.3. Comparison of TrpF structures

Two thermophilic bacterial TrpF structures are known: *Thermotoga maritima* TrpF (*Tm*TrpF; PDB entry 1lbm; Henn-Sax *et al.*, 2002) and *Thermus thermophilus* HB8 TrpF (*Tt*TrpF; PDB entry 1v5x; Taka *et al.*, 2005). The sequence identity between these structures and *Pf*TrpF is 29% for *Tm*TrpF and 27% for *Tt*TrpF, thus making it interesting to analyse whether these three proteins share the same mechanism of thermostability. We also compared the *Pf*TrpF structure with that of the analogous mesophilic protein from *E. coli*. In *E. coli* the last two steps of tryptophan synthesis occur on a dual-function enzyme and its C-terminal domain (*Ec*TrpF; residues 254–452 in PDB entry 1pii; Wilmanns *et al.*, 1992) has the same function as TrpF. *Ec*TrpF and *Pf*TrpF are 17% identical at the amino-acid level. These two thermophilic bacterial TrpFs and the corresponding domain of mesophilic *Ec*TrpF represent all the TrpF proteins that are currently available in the PDB. In the structure-based alignment, most of the barrel-strand sequences are better conserved than the helices, which is not surprising since they form the core of the structure (Fig. 3). The active-site loops which follow strands β5 and β6 and the α5 helix form the least conserved area. Both the loop following strand β5 and the following helix are a few amino acids longer in *Pf*TrpF than in the other enzymes. The *Ec*TrpF structure lacks helix α5; this α-helix is replaced by a loop structure connecting

Table 3
Potential salt bridges in *Pf*TrpF and their conservation in other analysed TrpF structures.

The connections are ordered based on the location of the acidic residue in the sequence and are classified according to which structural elements they connect and by further locating them to the barrel, helix or N- and C-terminal loops in the three-dimensional structure. The distances between residues are reported as the shortest distance between the side-chain O and N atoms.

Acidic residue	Basic residue	Connection	Distance (Å)	Location	Conservation
Glu13	Arg34	Helix–loop	3.2	C-loops	
Asp22	Lys204	Barrel–helix	2.6	N-loops	*Tm*TrpF
Glu39	Arg42	Helix–helix	3.1	α2	
Glu43	Lys40	Helix–helix	3.1	α2	
Asp88	Lys92	Helix–helix	3.6	α4	*Tt*TrpF†
Glu121	Arg117	Helix–helix	3.4	α5	
Asp128	Lys91	Barrel–helix	2.7	N-loops	
Asp180	Lys4	Barrel–barrel	2.8	β8–β1	*Ec*TrpF, *Tm*TrpF, *Tt*TrpF
(Glu186‡	Arg34)	Loop–loop		C-loops	*Ec*TrpF, *Tm*TrpF, *Tt*TrpF
Glu186	Lys191	Loop–loop	2.4	C-loops	*Ec*TrpF, *Tm*TrpF, *Tt*TrpF
Glu197	Arg201	Helix–helix	2.8	α8	
Glu198	Arg202	Helix–helix	2.7	α8	

† In *Tt*TrpF this interaction is between Glu and Arg residues. ‡ This interaction is not conserved in *Pf*TrpF, although the residues are.

the adjacent β-strands. The loop following strand β6 is only built in two of the structural models, *Tt*TrpF and *Ec*TrpF, making comparison of the β6 loop between the structures impossible.

The overall amino-acid composition of these four enzymes is somewhat different. Proline is the most rigid amino acid, whereas glycine, which lacks a side chain, can adopt different conformations more flexibly. The thermophilic *Tt*TrpF has the highest proportion of prolines in its sequence and an equal number of glycine residues (18), whereas the other two thermophilic enzymes also have an equal or almost equal number of prolines and glycines (8/11 in *Pf*TrpF and 11/11 in *Tm*TrpF; Table 2). However, the mesophilic *Ec*TrpF is quite distinct with respect to the number of glycine residues (21), which is more than four times higher than the number of prolines (five) (Table 2). Most of the conserved glycines are located in the loop area following the barrel strands in the active-site area, except for one located in the β2 strand of the barrel (Fig. 3).

3.4. Ionic interactions in the TrpF structures

The thermophilic TrpF structures contain more strong ion pairs, as they all contain 11–13 interactions involving charged Asp, Glu, Lys and Arg residues, whereas *Ec*TrpF only has eight interactions (Table 2). This is also reflected in the overall number of positively charged Arg and acidic Asp and Glu residues, which is higher in the thermophilic TrpF sequences than in *Ec*TrpF (Table 2). The thermophilic enzymes also have at least twice the number of acidic Glu/Asp residues compared with the neutral analogues Gln/Asn, whereas in *Ec*TrpF these are present in almost equal proportions (22 Glu/Asp residues and 21 Gln/Asn residues; Table 2).

The salt bridges are not well conserved within the structures that we have analysed. The only charged interactions that are conserved in all TrpFs are the ion pair between β1 and β8 that closes the barrel (between Asp180 and Lys4 in *Pf*TrpF) and the interaction stabilizing the loops close to the active site (Glu186–Lys191 in *Pf*TrpF) (Table 3). The interaction between Glu186 and Arg34 is absent in *Pf*TrpF as Arg34 forms an ion pair with Glu13 in the α1 helix. The Asp22–Lys204 interaction is only present in *Pf*TrpF and *Tm*TrpF. The rest of the salt bridges are unique to the individual structures (Table 3, Supplementary Table 1[1]).

Short-range local salt bridges connecting adjacent turns of an α-helix are typically observed in TrpF structures, although in *Ec*TrpF they only support two helices (Table 3, Supplementary Table 1[1]). *Tm*TrpF has surface salt bridges in five helices, whereas *Tt*TrpF has them in three helices. In addition (but only in *Tt*TrpF), there are also two salt bridges between two surface helices (Supplementary Table 1). *Pf*TrpF has a single salt bridge formed by α4 and α5, but it additionally contains paired ionic interactions that stabilize the surfaces of helices α2 and α8 (Table 3). These short-range interactions are most likely to contribute to the local stabilization of the helical structure and thus may reduce the tendency for unfolding. The most visible example of this is the α5 helix, which is long and well defined in *Pf*TrpF but has no secondary structure in *Ec*TrpF (Fig. 4).

3.5. Core packing and internal waters

The highly conserved barrel-forming residues indicate that the packing of the TrpF core barrel is highly similar in all structures (Fig. 3). None of the thermophilic structures have significant empty cavities in their core, while the mesophilic *Ec*TrpF has a few larger cavities close to the protein surface. Thermophilic TrpF structures indeed have better packing in the core compared with the mesophilic structure.

The number and identity of potential buried water-binding sites, however, is significantly different between the four TrpF structures. *Ec*TrpF includes 11 energetically favourable water molecules (*DOWSER* waters) per monomer, while the thermostable bacterial TrpF structures each contain nine *DOWSER* waters per monomer. The structure of *Pf*TrpF includes only four energetically favoured positions for buried water molecules (Fig. 4).

Most of the internal water molecules are located on the active-site face of the barrel. None of them are packed inside the barrel core of the protein. A few of the buried water binding sites are conserved in several structures, but only one is conserved in all four TrpF structures. This water site is embedded by conserved backbone hydrogen bonding from β2 and the loops following strands β1 and β2 in all structures (Wat1; Fig. 4).

Two other buried water molecules are conserved in the other three TrpF structures but not in *Pf*TrpF. In *Tt*TrpF, *Tm*TrpF and *Ec*TrpF, Wat2 (Fig. 4) is buried with conserved backbone hydrogen bonding from the loops following β1 and β8 and an aspartate residue in α1. In *Pf*TrpF the aspartate is replaced by Glu13 (Fig. 5), which stabilizes the above loops by

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: KW5046). Services for accessing this material are described at the back of the journal.

directly hydrogen bonding to the backbone N atoms of residues 9 and 191 and by forming additional hydrogen bonds to the side chain of Ser10 and a charged interaction with Arg34 (Fig. 5). Thus, the side chain of Glu13 replaces the interactions with Wat2 in the other structures. The loop following β8 is one residue shorter in *Pf*TrpF than in other structures (Fig. 3), thus also adopting a slightly different orientation in *Pf*TrpF (Fig. 5).

The other well conserved buried water, which is not present in *Pf*TrpF (Wat3 in *Tt*TrpF; Fig. 4), forms a bridge between the loops following β3 and β4. It is held in place by a β4-loop glutamate residue which is conserved in all of the other structures but not in *Pf*TrpF (Fig. 3). Instead, *Pf*TrpF uses Ser80 two residues earlier in the sequence, which hydrogen-bonds directly to the carbonyl O atom of Met58 in the β3 loop for local stabilization. As a consequence, the β3 loop also adopts a different conformation in *Pf*TrpF compared with the other three structures (Fig. 4). Again, a direct interaction between a protein residue and the backbone in *Pf*TrpF

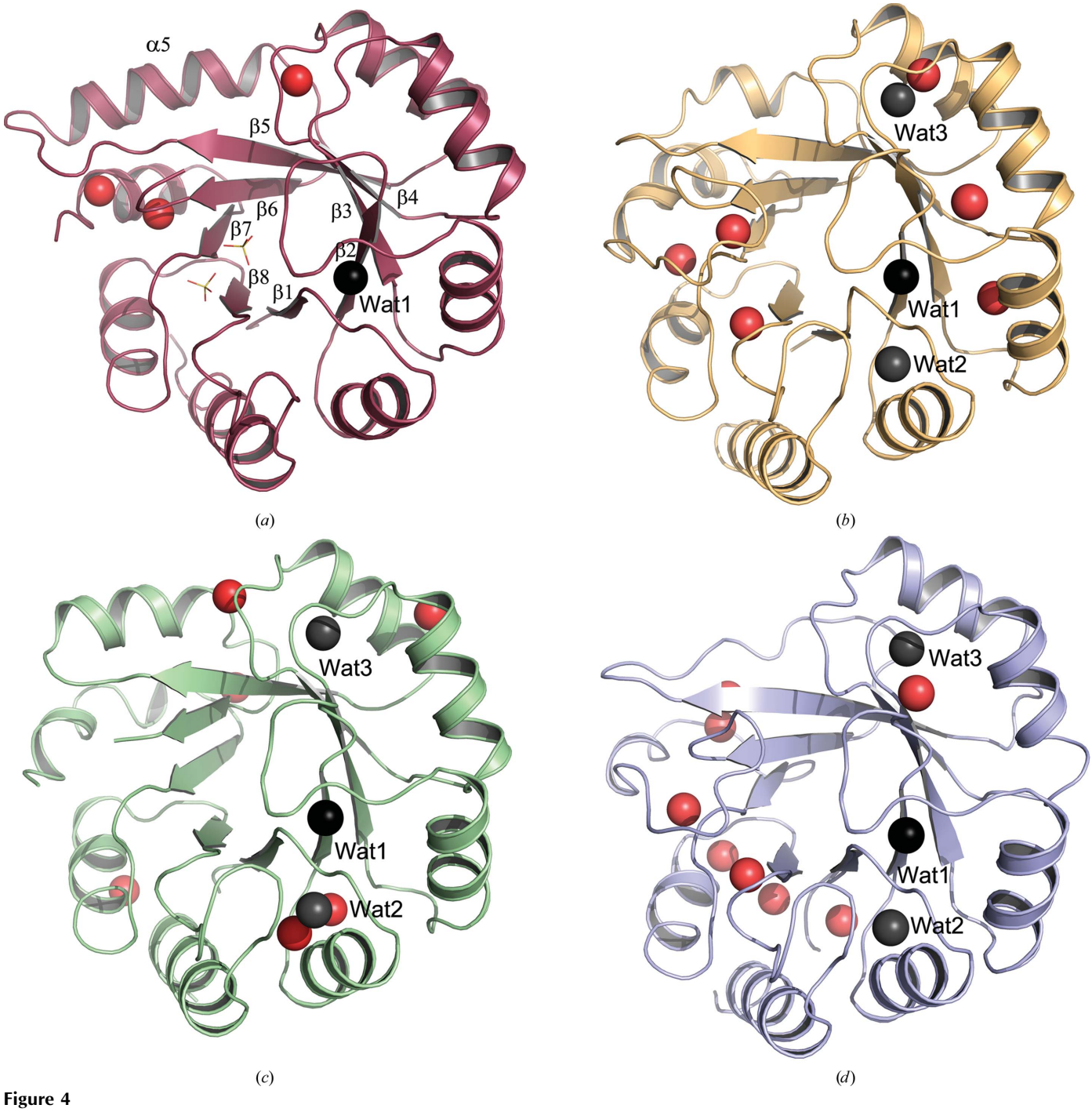

Figure 4
Buried *DOWSER* waters in TrpF structures. *Pf*TrpF (*a*) contains fewer *DOWSER* water sites than the other structures *Tt*TrpF (*b*), *Tm*TrpF (*c*) and *Ec*TrpF (*d*). Waters in unique buried sites are coloured red, waters in well conserved sites are coloured grey and the water-binding site conserved in all structures is coloured black. Other ligands in *Pf*TrpF and *Tm*TrpF are drawn as lines.

replaces the water-mediated stabilization in the bacterial TrpF structures.

4. Discussion

The TIM-barrel is by far the most common enzymatic fold; it has been estimated that 10% of all enzymes adopt this fold (Wierenga, 2001; Höcker *et al.*, 2001). The central β-barrel defines two distinct faces of the protein structure; the catalytic site is formed by the C-terminal loops emerging from the central β-barrel, while the N-terminal end of the barrel often has a role in protein stabilization (Höcker *et al.*, 2001). This is also evident in the TrpF family.

*Tm*TrpF and *Tt*TrpF are dimeric and the two monomers are interlocked by a hydrophobic loop at the N-terminal end of the barrel (Fig. 2*b*). In both cases low pH or deletion of this loop results in a monomeric protein which is far less stable (Taka *et al.*, 2005; Thoma *et al.*, 2000), suggesting that dimerization is obligatory for the thermostability of the bacterial TrpFs. However, the archaeal *Pf*TrpF and the two bacterial TrpFs differ in their oligomeric structures. *Pf*TrpF lacks the dimerization loop and is monomeric which, combined with its thermostability, makes *Pf*TrpF an interesting enzyme-engineering template for industrial applications. How is this monomeric TIM-barrel structure stabilized?

Protein stability is thermodynamically defined as the free-energy difference between the folded and unfolded states (ΔG). Thus, stability can be increased by stabilizing the folded state, for example by increasing the number of salt bridges or hydrogen bonds. Another way is to destabilize the unfolded state, *e.g.* by reducing ΔS (Razvi & Scholtz, 2006). Adjustment of the entropic loss upon folding is increasingly important at high temperatures, as ΔS is the temperature-sensitive term in the Gibbs free-energy calculation (Razvi & Scholtz, 2006). Higher stability can also be achieved by changing the kinetic stability, which affects the rate of protein unfolding (Vieille & Zeikus, 2001).

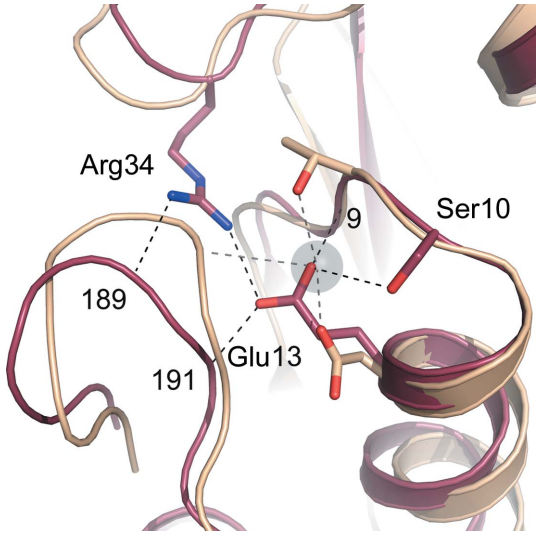

Figure 5
Replacement of a conserved water-binding site in *Pf*TrpF. (a) Glu13 in *Pf*TrpF (red structure) replaces Asp13 in *Tt*TrpF (light brown structure), which coordinates a water molecule (Wat2 in *Tt*TrpF; dark grey sphere) in the other three structures. In *Pf*TrpF the water coordination is replaced by hydrogen bonding from Glu13 and a charged interaction with Arg34. Residues which only participate in backbone hydrogen bonding are labelled with the residue number only.

The most apparent way to affect the free-energy difference between the folded and the unfolded states is to increase the number of ionic interactions. The *Pf*TrpF structure has 11 salt-bridge interactions, which is significantly more than in *Ec*TrpF but not substantially different from the other two thermostable TrpFs (Table 3). The most conserved interactions combine residues distant in sequence and connect the first and last strands of the barrel (Asp180–Lys4) or β2 to the end of α8 (Asp22–Lys204), thus most likely stabilizing the closure of the core barrel. The other individual ion pairs are not conserved. However, ∼30% of charged core residues involved in ionic interactions additionally participate in side-chain–main-chain interactions (32% in *Pf*TrpF, 31% in *Ec*TrpF, 28% in *Tm*TrpF and 22% in *Tt*TrpF), thus contributing to the stability of the fold. In the TrpF structures the surface helices typically contain short-range interlinking ion pairs within the α-helices, which favour these units as integrated building blocks for the TIM-barrel fold, but as these mechanisms are common to all thermophilic TrpFs they are not sufficient to explain the hyperthermophilic nature of the *Pf*TrpF monomer.

Some of the adjustment of the entropic term is possible to analyse at the sequence level. Increasing the number of proline residues and reducing the number of glycines will restrict the conformational flexibility of the unfolded state and thus reduce the entropic loss upon folding. This mechanism may not be ideal for enzymes, which often require flexibility for their function, and indeed is mostly used by structural proteins, which tolerate rigid conformations better (Razvi & Scholtz, 2006). Amongst the bacterial TrpFs, the two thermophilic enzymes contain a larger proportion of prolines in comparison to the mesophilic *Ec*TrpF. However, the hyperthermophilic archaeal *Pf*TrpF has almost as low a proline content as *Ec*TrpF (Table 2), although by definition it is more stable than *Ec*TrpF. Thus, a proline-rich sequence does not appear to be the only viable solution for creating a thermostable TIM-barrel enzyme.

Buried water molecules may significantly contribute to protein stability by filling up possible cavities formed during protein folding. Buried waters provide hydrogen-bonding networks, yet allow flexible interaction with the surrounding solvent during protein function. However, as water molecules are not covalently attached to the protein chain their contribution to stabilization is limited. Psychrophilic enzymes tend to include more water-filled cavities than other proteins (Paredes *et al.*, 2011), but at elevated temperatures thermal motion of the water molecules can be significantly destabilizing.

The most evident structural difference between *Pf*TrpF and the dimeric bacterial TrpFs is the way that they incorporate water molecules. Three *DOWSER* water molecules are conserved among *Tm*TrpF, *Tt*TrpF and *Ec*TrpF (Fig. 4), which

points to a key role of these structural waters in the bacterial TrpFs. However, only one of these conserved waters is present in *Pf*TrpF, which has fewer buried water binding sites in total. One of the conserved water molecules among *Tm*TrpF, *Tt*TrpF and *Ec*TrpF was replaced by a side-chain interaction of Glu13 in *Pf*TrpF. This residue is highly conserved in archaeal TrpF sequences, which suggests that archaeal TrpFs might generally use this mechanism for improving their thermal stability.

The loops which cover the buried waters in the bacterial TrpFs are shorter in *Pf*TrpF. Instead of water molecules, *Pf*TrpF uses the intrinsic properties of the polypeptide chain for local and global stabilization of the corresponding structure (Fig. 5). Direct protein side-chain–main-chain interactions in *Pf*TrpF replace the structural water molecules present in the other TrpFs, thus providing a maximal gain in protein stabilization but minimizing the entropic cost of burying structural waters upon folding.

Our comparative structural analysis indicates that reduction of disorder is the main mechanism that *P. furiosus* uses to ensure protein function at elevated temperatures. This creates a thermodynamically stable enzyme fold which is supported by charged interactions. In addition, *Pf*TrpF contains fewer structural water molecules by having a unique solution of replacing structural water molecules by side-chain–main-chain interactions. At elevated temperatures, entropy and high thermal motion of waters makes them destabilizing. Therefore, replacement of structural water molecules by direct side-chain–main-chain interactions in *Pf*TrpF could be an effective way of stabilizing the protein structure at higher temperatures. By combining these structural mechanisms, the functional monomeric TIM-barrel enzyme from *P. furiosus* remains stable at the temperatures of the geothermally heated marine sediments from which this archaeal species was isolated.

We thank MSc Kimmo Heinämäki for help with the preliminary crystallization experiments, Dr Jan Kadlec for help in data collection at ESRF beamline ID14-4 and Dr Anssi Malinen for helpful comments during manuscript preparation. This project is supported by Biocenter Finland and the Academy of Finland (grant No. 131413 to HI).

References

Balamurugan, B. *et al.* (2007). *J. Appl. Cryst.* **40**, 773–777.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
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.
Creighton, T. E. (1994). *Proteins: Structures and Molecular Properties.* New York: W. H. Freeman & Co.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Henn-Sax, M., Thoma, R., Schmidt, S., Hennig, M., Kirschner, K. & Sterner, R. (2002). *Biochemistry*, **41**, 12032–12042.
Höcker, B., Jürgens, C., Wilmanns, M. & Sterner, R. (2001). *Curr. Opin. Biotechnol.* **12**, 376–381.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kleywegt, G. J. & Jones, T. A. (1994). *Acta Cryst.* D**50**, 178–185.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. (2005). *Nucleic Acids Res.* **33**, D266–D268.
Luginbühl, P., Pervushin, K. V., Iwai, H. & Wüthrich, K. (1997). *Biochemistry*, **36**, 7305–7312.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Orekhov, V. Y., Nolde, D. E., Golovanov, A. P., Korzhnev, D. M. & Arseniev, A. S. (1995). *Appl. Magn. Reson.* **9**, 581–588.
Ortega, A. & García de la Torre, J. (2005). *J. Am. Chem. Soc.* **127**, 12764–12765.
Paredes, D. I., Watters, K., Pitman, D. J., Bystroff, C. & Dordick, J. S. (2011). *BMC Struct. Biol.* **11**, 42.
Razvi, A. & Scholtz, J. M. (2006). *Protein Sci.* **15**, 1569–1578.
Reddy, T. & Rainey, J. K. (2010). *Biochem. Cell Biol.* **88**, 131–142.
Röthlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S. & Baker, D. (2008). *Nature (London)*, **453**, 190–195.
Sterner, R., Kleemann, G. R., Szadkowski, H., Lustig, A., Hennig, M. & Kirschner, K. (1996). *Protein Sci.* **5**, 2000–2008.
Taka, J., Ogasahara, K., Jeyakanthan, J., Kunishima, N., Kuroishi, C., Sugahara, M., Yokoyama, S. & Yutani, K. (2005). *J. Biochem.* **137**, 569–578.
Thoma, R., Hennig, M., Sterner, R. & Kirschner, K. (2000). *Structure*, **8**, 265–276.
Vieille, C. & Zeikus, G. J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 1–43.
Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. (2005). *Proteins*, **59**, 687–696.
Wierenga, R. K. (2001). *FEBS Lett.* **492**, 193–198.
Wilmanns, M., Priestle, J. P., Niermann, T. & Jansonius, J. N. (1992). *J. Mol. Biol.* **223**, 477–507.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zhang, L. & Hermans, J. (1996). *Proteins*, **24**, 433–438.