Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of HisF, a histidine biosynthetic protein from *Pyrobaculum aerophilum*

Mark J. Banfield,[a,b] J. Shaun Lott,[a] Vickery L. Arcus,[a] Andrew A. McCarthy[a] and Edward N. Baker[a,c]*

[a]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand, [b]Department of Biochemistry, University of Bristol, Bristol BS8 1TD, England, and [c]Department of Chemistry, University of Auckland, Private Bag 92019, Auckland, New Zealand

Correspondence e-mail: ted.baker@auckland.ac.nz

Received 12 June 2001
Accepted 24 July 2001

PDB Reference: HisF, 1h5y.

HisF (imidazole glycerol phosphate synthase) is an important branch-point enzyme in the histidine biosynthetic pathway of microorganisms. Because of its potential relevance for structure-based drug design, the crystal structure of HisF from the hyperthermophilic archaeon *Pyrobaculum aerophilum* has been determined. The structure was determined by molecular replacement and refined at 2.0 Å resolution to a crystallographic *R* factor of 20.6% and a free *R* of 22.7%. The structure adopts a classic $(\beta/\alpha)_8$ barrel fold and has networks of surface salt bridges that may contribute to thermostability. The active site is marked out by the presence of two bound phosphate ions and two glycerol molecules that delineate a long groove at one end of the $(\beta/\alpha)_8$ barrel. The two phosphate ions, 17 Å apart, are bound to sequence-conserved structural motifs that seem likely to provide much of the specificity for the two phosphate groups of the HisF substrate. The two glycerol molecules bind in the vicinity of other sequence-conserved residues that are likely to be involved in binding and/or catalysis. Comparisons with the homologous HisF from *Thermatoga maritima* reveal a displaced loop that may serve as a lid over the active site.

1. Introduction

Histidine is a so-called essential amino acid because humans and other higher eukaryotes lack the biosynthetic apparatus for its manufacture. On the other hand, the genes that code for the enzymes of the histidine biosynthetic pathway are widely present in bacteria and these enzymes are therefore attractive targets for the design of novel antibacterial drugs. In *Escherichia coli*, the biosynthesis of histidine takes place in a series of ten enzymatic reactions performed by the products of eight genes labelled *hisA* to *hisI* (Alifano *et al.*, 1996). By inference, a similar biosynthetic pathway is known or expected to operate throughout eubacteria, in some lower eukaryotes and possibly in archaea. The *his* genes also provide one of the best examples of the grouping of genes of related functions into operons and studies of the regulation of histidine biosynthetic enzymes (*e.g.* Ames & Garry, 1959) led to the formal enunciation of the operon concept (Jacob & Monod, 1961). Some of the genes and their products also give clear evidence of having arisen by gene duplication and subsequent elaboration of function; for example, the *hisA* and *hisF* gene products catalyse adjacent steps in the biosynthetic pathway (Alifano *et al.*, 1996) and have been shown to have homologous sequences, with ∼25% sequence identity (Fani *et al.*, 1994; Thoma *et al.*, 1998) and very similar three-dimensional structures (Lang *et al.*, 2000).

In *E. coli*, HisF forms a heterodimeric complex with HisH that catalyses the ammonolytic splitting of *N*′-[(5′-ribulosyl)-formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (5′-PRFAR) to imidazole glycerol phosphate (IGP) and 5-aminoimidazole-4-carboxamide-1-β-D-ribofuranosyl 5′-monophosphate (AICAR) (Klem & Davisson, 1993; Rieder *et al.*, 1994; Alifano *et al.*, 1996). In this complex, referred to as IGP synthase (Klem & Davisson, 1993), the enzymatic reaction occurs on the HisF subunit, with HisH supplying ammonia derived from glutamine (see Fig. 1 for a reaction scheme). Functionally similar heterodimeric complexes include the PabB–PabA complex used in folate biosynthesis in *E. coli* and the TrpG–TrpE complex used in tryptophan biosynthesis in *Serratia marcescans* (Green & Nichols, 1991; Rieder *et al.*, 1994). The HisF–HisH complex is an important branch-point enzyme, as the IGP is used in the continued synthesis of histidine and the second product AICAR is used for *de novo* purine biosynthesis.

Our interest in HisF arose in the context of a structural genomics initiative focused on the hyperthermophilic archaeon *P. aerophilum*. We sought to identify genes present in *P. aerophilum* that are also present, with a significant degree of sequence identity, in pathogenic organisms such as *Mycobacterium tuberculosis*. Amongst these is the *P. aerophilum* gene pag5_989, the protein product of which has approximately 56% sequence identity with the protein encoded by the *hisF* gene (Rv1605) in the *M. tuberculosis* genome (Cole *et al.*, 1998). Since this work began, the structure of HisF from another thermophile, *Thermatoga maritima*, has been independently determined (Lang *et al.*, 2000). The structure of this protein and subsequent dissection experiments on the $(\beta/\alpha)_8$ barrel (Höcker *et al.*, 2001) suggest it is made up of two subdomains and that HisF evolved by tandem duplication and fusion from an ancestral half-barrel. *P. aerophilum* HisF shares 55% sequence identity with the *T. maritima* protein.

Here, we present the structure of HisF from *P. aerophilum* at 2.0 Å resolution, compare it with that of HisF from *T. maritima* and identify the likely substrate-binding site based on sequence conservation and the presence of two bound phosphate ions and two glycerol molecules in the putative active site.

2. Materials and methods

2.1. Cloning and expression

The DNA encoding the *P. aerophilum* HisF protein was obtained by PCR from genomic DNA. The ORF was located in the genomic sequence and primers were designed to amplify the relevant region and incorporate the desired restriction enzyme sites for subsequent cloning into the pET28a expression vector (Novagen). This led to the design of a 30 base-pair 5′ (left) primer that had three mismatched bases compared with the genomic sequence to create an *Nde*I site (5′-TTAACTTGTG**CATATG**GCCTTGAGGATAAT-3′) and a 26 base-pair 3′ (right) primer that spanned a *Sal*I site present in the genome, approximately 40 bases downstream of the stop codon of the ORF (5′-TTTGCACGTC**GTCGAC**GATTTT-CCAC-3′). The resulting ∼900 base-pair PCR fragment was cloned into pET28a using the *Nde*I and *Sal*I sites, then transformed first into the *E. coli* cloning strain DH5α and subsequently into the expression strain BL21(DE3). The BL21(DE3) cells were pre-transformed with a plasmid (pRI952) that encodes tRNAs to rare codons in *E. coli* in order to enhance expression levels. Small-scale tests confirmed that the protein was expressed in soluble form.

2.2. Purification and crystallization

The HisF protein was purified from bacterial cultures (2–4 l) by metal chelation chromatography (using the poly-His tag on the N-terminus) followed by gel filtration. Harvested cells were lysed by sonication in loading buffer (50 m*M* HEPES, 150 m*M* NaCl, 20 m*M* imidazole pH 7.5), then centrifuged at 13 000*g* to pellet cell debris. The resulting supernatant was loaded onto a 5 ml His-trap column (Pharmacia). The column was washed with 5–10 column volumes of loading buffer before eluting the protein with a step gradient of loading and elution buffer (50 m*M* HEPES, 150 m*M* NaCl, 1 *M* imidazole pH 7.5). Fractions containing HisF were pooled and concentrated before loading onto a Superdex75 gel-filtration column (Pharmacia) equilibrated with 50 m*M* Tris, 150 m*M* NaCl pH 7.5. The protein eluted as a single peak.

For crystallization, the protein solution was concentrated to 10 mg ml^{-1}. Initial crystallization conditions were obtained from commercially available crystallization screens. Optimization of these conditions produced diffraction-quality crystals from 100 m*M* sodium citrate pH 5.6, 0.9–1.0 *M* ammonium dihydrogen phosphate. The crystals formed were perfect hexagonal rods with tapered ends. Interestingly, these crystallization conditions are virtually identical to those found independently to crystallize *T. maritima* HisF (Thoma *et al.*, 1999), although different crystal lattices were obtained.

2.3. Data collection

Prior to data collection, the crystals were stabilized in 100 m*M* sodium citrate pH 5.6, 1.2 *M* ammonium dihydrogen phosphate and then washed in progressive steps of increasing glycerol concentration (5% steps for 10 min) to a final cryoprotective concentration of 25% glycerol. Crystals were then flash-frozen in a liquid-nitrogen cold stream at 100 K.

An initial diffraction data set was collected on a MAR345 image plate mounted on a Rigaku RU-300 rotating-anode generator equipped with focusing mirrors. The crystals diffracted to 3.1 Å resolution and had unit-cell parameters $a = b = 105.74$, $c = 141.33$ Å. Diffraction symmetry was hexagonal, in accord with the crystal morphology. Despite the low resolution, the data were of high quality, as judged by crystal mosaicity, redundancy, completeness and $\langle I\rangle/\sigma(I)$ (Table 1) and these data were used in the initial structure solution (see below). A second data set was subsequently collected from a similarly flash-frozen crystal at the Stanford Synchrotron Radiation Laboratory (SSRL). This data set was collected with an ADSC Quantum-4 CCD detector on

Table 1
Data-collection statistics for the crystals of *P. aerophilum* HisF.

	Home data	SSRL data
Unit-cell parameters (Å)	$a = b = 105.74$, $c = 141.33$	$a = b = 105.90$, $c = 141.83$
Space group	$P6_1$	$P6_1$
Wavelength (Å)	1.54	1.00
Resolution range (Å)	35–3.1 (3.21–3.1)	25–2.0 (2.09–2.0)
No. of unique reflections	16233	59658
Redundancy	4.5 (4.4)	3.7 (2.4)
Completeness (%)	99.7 (99.4)	98.3 (99.9)
R_{merge}(%)	9.4 (27.7)	5.3 (20.8)
$\langle I\rangle/\sigma(I)$	15 (4.4)	25.1 (4.7)

Table 2
Refinement statistics for the final model of *P. aerophilum* HisF.

Resolution (Å)	25–2.0 (2.09–2.0)
R_{cryst} (%)	20.6 (24.0)
R_{free} (%)	22.7 (29.0)
R.m.s.d. bond lengths (Å)	0.019
R.m.s.d. bond angles (°)	1.70
No. of non-H protein atoms	3780
No. of non-H solvent atoms	313 (4 phosphate ions, 4 glycerol and 269 water molecules)
Overall *B* value (Å^2)	41.7
Main chain	40.4
Side chain	43.0
Solvent residues	46.4

beamline PX9-2 (with radiation of wavelength 1.00 Å). Diffraction data extended to 2.0 Å resolution and this data set was used to complete the refinement of the HisF model. All data processing was carried out with the *HKL* suite (Otwinowski & Minor, 1997).

2.4. Structure determination and refinement

The structure was solved by molecular replacement using *AMoRe* (Navaza, 1994) as implemented in the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) with the structure of the *T. maritima* HisF as a search model (PDB code 1thf). Translation searches confirmed the space group to be $P6_1$. The solution also confirmed that there were two molecules in the asymmetric unit, as was suspected from the volume of the unit cell.

The structure was refined with *REFMAC*5 (version 5.0.26; Murshudov *et al.*, 1997), initially against the home-laboratory data set, using all data in the resolution range 35–3.1 Å. A randomly chosen 5% of reflections were flagged for free *R* calculation (Brünger, 1992). The first round of refinement lowered the crystallographic *R* factor (R_{cryst}) to 39.3% and the free *R* factor (R_{free}) to 44.3% with tight non-crystallographic symmetry (NCS) restraints. At this stage, the model still comprised the *T. maritima* HisF sequence. The model was mutated to the *P. aerophilum* sequence during the first rebuild. Five iterative cycles of rebuilding with *O* (Jones *et al.*, 1991) and refinement with *REFMAC*5 lowered R_{cryst} and R_{free} to final values of 21.1 and 25.8%, with a root-mean-square (r.m.s.) deviation from standard values of 0.013 Å for bond lengths and 1.6° for bond angles. Medium NCS restraints were retained on the two molecules throughout refinement, except for the final stages where they were released. The refinement was then completed using the 2.0 Å resolution SSRL data set. The same set of free *R* reflections was retained as for the lower resolution refinement, with further randomly selected higher resolution reflections added. Data in the resolution range 25–2.0 Å were used. No NCS restraints were applied in this latter phase of refinement. The final values of R_{cryst} and R_{free} were 20.6 and 22.7%, respectively, with the r.m.s. deviations from standard values being 0.019 Å for bond lengths and 1.7° for bond angles. Refinement statistics are summarized in Table 2 and an example of the final electron density is shown in Fig. 2. Analysis with *PROCHECK* (Laskowski *et al.*, 1993) showed that 95.0% of residues were in the most favoured regions of the Ramachandran plot, with no residues in disallowed regions.

The final model comprises residues Ala4–Ile253 of the *P. aerophilum* HisF polypeptide, for both molecules in the asymmetric unit (all residues are numbered as shown in Fig. 3). Also present are two additional residues on the N-terminus for molecule *A* (His2, Met3) and three residues on the N-terminus for molecule *B* (labelled Ser1, His2 and Met3) that are cloning artefacts derived from the poly His tag, but are observed in the electron density. Residues 1–3 of the intact *P. aerophilum* sequence were not included in the DNA construct that was cloned and expressed. As judged from the position of the N-terminus in the structure, it is unlikely the absence of these residues affects the activity or conformation of the protein. The side chains of three residues from molecule *A* (Arg30, Arg121 and Asp222) have been modelled with alternate conformations, as have two side chains

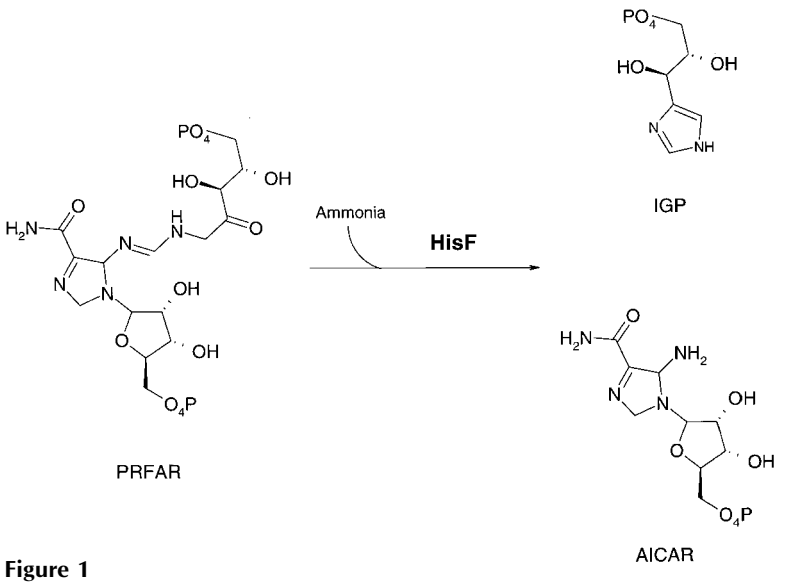

Figure 1
Reaction scheme for the cleavage reaction catalysed by HisF (ammonia donated by the HisH glutaminase activity). See text for abbreviations.

from molecule *B* (Arg30 and Asp222). Several residues have no interpretable electron density for some or all of their side-chain atoms; these include Lys18, Glu60, Arg97 and Arg198 from molecule *A*, and Lys18, Glu60, Arg97, Arg111, Arg121 and Arg198 from molecule *B*. Apart from the slight differences in these mobile side chains, the two independent molecules in the asymmetric unit have essentially identical structures, but do not form any dimer-like association. Each molecule also has two bound phosphate ions and two bound glycerol molecules.

3. Results and discussion

3.1. Sequence comparison

Owing to its central role in histidine biosynthesis, it is not surprising that the HisF protein is highly conserved in organisms that synthesize this amino acid from precursor molecules. A sequence alignment for a selection of HisF proteins from bacteria and archea, including mesophiles and hyperthermophiles, is shown in Fig. 3. Apart from variations at the N- and C-termini, the only significant insertions or deletions in the various HisF polypeptides are a two amino-acid insertion in the *P. aerophilum* HisF following residue 15 and insertions/deletions in a region following residue 136 (*P. aerophilum* numbering). Both of these regions are within loops at one end of the $(\beta/\alpha)_8$ barrel fold adopted by the HisF proteins (see later) and may play an important role in either catalysis or protein–protein interactions.

3.2. Overall structure of *P. aerophilum* HisF and comparison with its *T. maritima* homologue

P. aerophilum HisF adopts an archetypal $(\beta/\alpha)_8$-barrel (TIM barrel) structure (Fig. 4*a*). As might be expected from the observed level of sequence identity (55%), the overall fold of *P. aerophilum* HisF is similar to the *T. maritima* homologue; they can be superimposed with an r.m.s. difference in atomic positions of 0.99 Å (for 229 equivalent C^{α} atoms). There are only two regions of the polypeptide chain that differ significantly in conformation: the N-terminal 2–3 residues and the long loop between $\beta1$ and $\alpha1$ (residues 12–31), where a completely different conformation is adopted (Fig. 4*b*). In *P. aerophilum* HisF, this loop has a two-residue insertion compared with the *T. maritima* protein. The observed rearrangement is potentially very significant, as in *T. maritima* HisF the loop is positioned well away from the centre of the $(\beta/\alpha)_8$ barrel, leaving clear uninhibited access to the presumed active site, but in the *P. aerophilum* structure this loop bends over the centre of the barrel and Lys22 projects across the central cavity where it binds to a glycerol molecule at the entry to the active site. The loop partially blocks the access to the active site from bulk solvent and may serve as a moveable lid during substrate binding and release. If this is so, the most likely scenario is that the PRFAR substrate binds first, followed by movement of the loop to partially enclose it. Depending on where the HisH component of the heterodimer binds, movement of the loop could also facilitate the adoption of a productive binding mode for HisH once the PRFAR substrate is in place.

The insertions and deletions that follow residue 136 (*P. aerophilum* numbering) in many members of the HisF family seem much less likely to affect the catalytic site. These changes are located in a surface loop connecting $\beta5$ and $\beta5'$ on the outer surface of the C-terminal face of the $(\beta/\alpha)_8$ barrel. In the structures of *P. aerophilum* and *T. maritima* HisF, which share the same number of amino acids in this region, the loops have the same conformation. As discussed in the introduction, however, the *in vivo* mechanism of HisF involves the formation of a heterodimeric complex with HisH. It is thus conceivable that this loop may be involved in mediating species-specific interaction with HisH.

3.3. Active site

As in other $(\beta/\alpha)_8$-barrel enzymes, the active site of the enzyme is presumed to be located at the end of the $(\beta/\alpha)_8$ barrel where loops connect the C-terminal ends of the β-strands to the helix N-termini. These loops provide the variations that allow different catalytic activities and substrate-binding ability to be supported in different enzymes, using the same protein framework. For HisF, this conclusion is supported by the location at this end of the barrel of a cluster of invariant residues that have been identified in multiple sequence alignments (Thoma *et al.*, 1998).

Further confirmation of the active site is given by the presence of two phosphate ions that are bound in this region in both the *P. aerophilum* and *T. maritima* structures. In *P. aerophilum* HisF, the two phosphate ions are bound 17 Å apart at either end of a long groove formed by the connecting loops at this end of the barrel (Fig. 5). These phosphate ions are presumably derived from the crystallization media, but almost certainly mimic the positions of the two phosphoryl groups in the HisF substrate. Phosphate ion *A* is bound in a pocket formed by loops belonging to the N-terminal half of the barrel, primarily the $\beta3$–$\alpha3$ connecting loop (residues 83–86) and the $\beta4$–$\alpha4'$ connecting loop (residues 106–109). These two loops bind the phosphate ion through three peptide NH groups and the side chain OH of Thr107 and there are additional water bridges to the loop connecting $\beta5'$ and $\beta5''$. A further loop, $\beta2$–$\alpha2$ (residues 51–60), covers the phosphate site from above so that it is largely enclosed by protein. The second phosphate ion *B* is similarly bound in a pocket lined by residues from the $\beta7$–$\alpha7$ loop (residues 204–208) and the loop between $\beta8$ and $\alpha8'$ (residues 227–229); the latter two loops are the C-terminal half equivalents of the $\beta3$–$\alpha3$ and $\beta4$–$\alpha4'$ loops. It is also partially covered by residues from the $\beta6$–$\alpha6$ connection (residues 176–181). The binding site involves four peptide NH groups, the side chain of Ser228 and several tightly bound water molecules.

The sequences of these phosphate-recognition loops are highly conserved, G(A)GGV(I) for 83–86, NT(S)*x*A for 106–

109, SG(S)GA(V)G for 204–208 and AS(G)I(V) for 227–229 (Fig. 2). The sites are the same as in *T. maritima* HisF (Fig. 4*b*) and the sequence conservation across the entire family implies that these are critical motifs for the recognition and binding of the bisphosphorylated substrate. Phosphate binding will be enhanced by the fact that both phosphate sites are at the N-termini of α-helices, helices $\alpha4'$ (107–113) and $\alpha8'$ (228–233) (see Figs. 4 and 5).

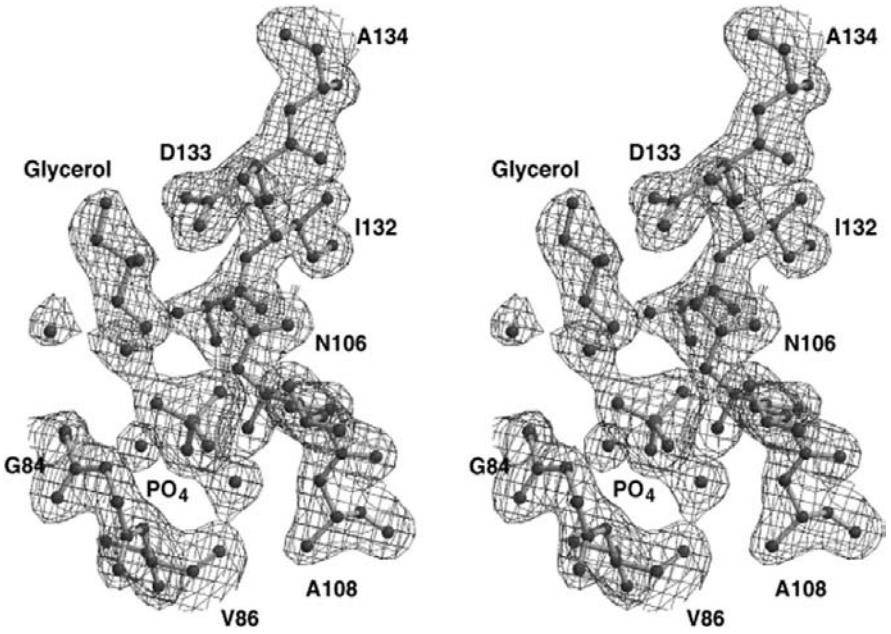

Figure 2
An example of the observed electron density and corresponding residues from the final model of *P. aerophilum* HisF. The electron density is calculated from σ_A-weighted $2|F_{obs}| - |F_{calc}|$ coefficients and is contoured at 1.2σ. Residues shown are Gly84–Val86, Asn106–Ala108, Ala131–Ala134, along with the phosphate that occupies the *A* site, the tightly bound glycerol molecule and four water molecules. The figure was prepared with *BOBSCRIPT* (Esnouf, 1997) and *RASTER3D* (Merritt & Bacon, 1997).

In our structure we also have two molecules of glycerol bound within each monomer that were not observed in *T. maritima* HisF. These glycerol molecules are presumed to be derived from the cryoprotectant solutions that the crystals were soaked in prior to flash-freezing. They are well ordered, with occupancies of 100%. Both are bound in the active-site groove and are likely to delineate parts of the binding site of the HisF substrate. One glycerol molecule is deeply bound, with its terminal OH approximately 4.5 Å from phosphate *A*, and is hydrogen bonded through its O2 and O3 atoms to the OD1 and OD2 atoms of Asp133 (Fig. 5). Asp133 is invariant in all HisF proteins and is assumed to be an essential catalytic residue. We propose that the interaction between the glycerol and Asp133 is functionally significant and mimics the way the glycero-phosphate part of the PRFAR substrate binds to the protein. The second of the glycerol molecules is bound nearer the surface of the active

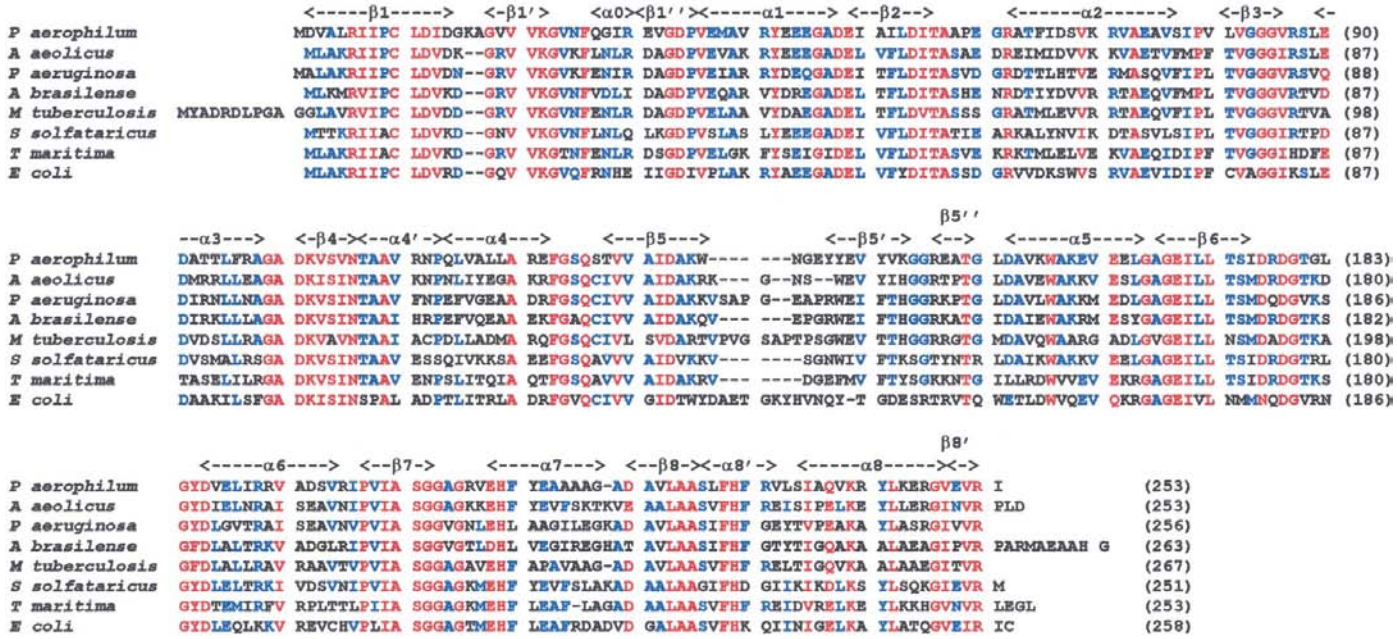

Figure 3
Sequence alignment of a selection of HisF proteins based on the alignment of the top ten hits to *P. aerophilum* HisF identified from a FASTA3 search of the SWISS-PROT database. Residues in red are conserved in >90% of the sequences aligned, residues in blue are conserved in >50% (I/V, L/M, F/Y, D/E/N/Q are considered conserved). The alignment contains four thermophilic (*Pyrobaculum aerophilum*, *Aquifex aeolicus*, *Sulfolobus solfataricus*, *Thermotoga maritima*) and four mesophilic (*Pseudomonas aeruginosa*, *Azospirillium brasilense*, *Mycobacterium tuberculosis*, *Escherichia coli*) homologues. Figure prepared with *MULTALIN* (Corpet, 1988).

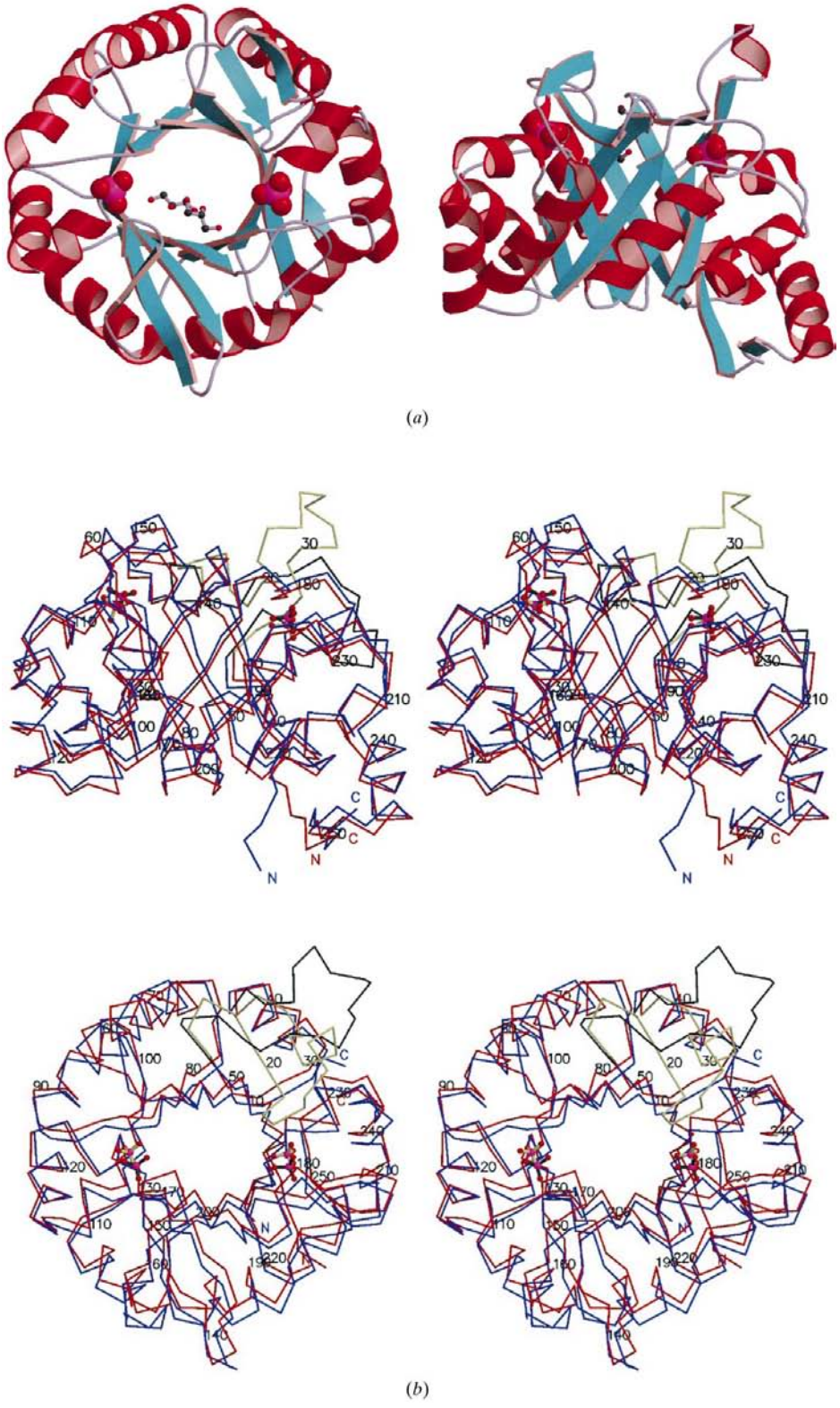

Figure 4
(*a*) The overall structure of *P. aerophilum* HisF displayed in two approximately orthogonal views. The two phosphate ions and two glycerol molecules observed in the structure are also shown in CPK/ball-and-stick representation. (*b*) Two approximately orthogonal stereoviews showing an overlay of the structures of *P. aerophilum* (red) and *T. maritima* (blue) HisF. The loop between β1 and α1 that adopts a different conformation in the two proteins is highlighted (*P. aerophilum*, khaki; *T. maritima*, dark grey). Also shown are the phosphate ions in each structure (sticks shown in the same colour scheme as the loops) and every tenth amino acid in the *P. aerophilum* structure is numbered. The figure was prepared with *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D*.

site, where the HisH subunit may dock. It is within hydrogen-bonding distance of the first glycerol and is also hydrogen bonded to the side chains of two conserved side chains, Lys22 and Asp179 (Fig. 5). This glycerol may mark the site of the amidotransferase reaction, in which glutamine provided by HisH is used to cleave the PRFAR substrate and complete the synthesis of the imidazole glycerol phosphate moiety.

The *T. maritima* HisF $(\beta/\alpha)_8$ barrel is thought to have evolved by tandem duplication from an ancestral half-barrel, with both structural and sequence homology between the N- and C-terminal halves of the barrel. For *P. aerophilum* HisF, the sequence identity between the two halves is ~25% and the structure reveals a corresponding twofold internal structural homology. Evolution subsequent to the original duplication has, however, adapted the protein to its asymmetric substrate. It seems highly likely from the *P. aerophilum* structure that the substrate binds in an orientation that places the glycero-phosphate part of PRFAR at a position where the glycero moiety forms interactions with Asp133 and the phosphate binds to residues in the pocket *A* described above. When considering the structure as a duplicated half-barrel there is no evidence from the electron density to suggest that a glycerol molecule is bound to Asp12, a second essential catalytic residue that occupies the equivalent position to Asp133 in the second half-barrel. This further supports our association of the glycero-phosphate moiety of PRFAR with Asp133 and phosphate site *A*. The logical extension of this is that Asp12 is likely to bind the 2′- and 3′-OH groups of the PRFAR ribose moiety; this would then locate the second phosphate of PRFAR to the phosphate-binding site *B* observed in the *P. aerophilum* and *T. maritima* structures.

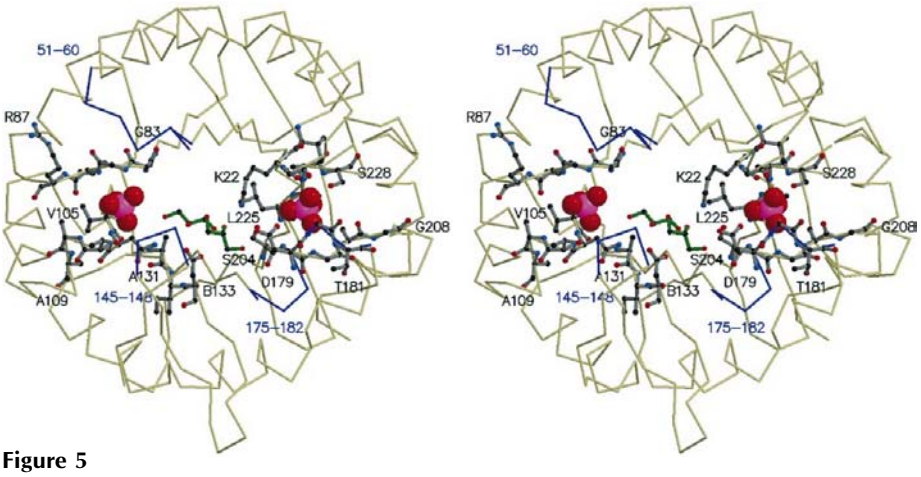

Figure 5
Overall view of the putative active site, including the phosphate (CPK representation) and glycerol molecules (sticks in green), along with relevant amino acids as discussed in the text (21–23, 83–87, 105–109, 131–133, 179–181, 204–208, 225–228); the loops that enclose the phosphates are shown in blue. The figure was prepared with *MOLSCRIPT* and *RASTER3D*.

3.4. Thermostability

P. aerophilum is a hyperthermophilic archaeum that grows optimally at 373 K (Volki *et al.*, 1996). Although no structure is yet available for HisF from any mesophilic organism, the availability of the *P. aerophilum* HisF structure, coupled with the sequence data available for HisF from many organisms, allows us to seek structural features that may confer extreme thermostability on the *P. aerophilum* enzyme. Factors that may contribute to enhanced thermostability include, for example, reduced numbers of amino acids that are prone to deamidation (Asn, Gln) or oxidation (Cys, Met) and a variety of structural strategies including increased numbers of salt bridges, hydrogen bonds or hydrophobic interactions, shortened loops or chain termini, reduced hydrophobic surface, metal binding and enhanced subunit associations. These factors have different levels of prominence in different proteins (Knapp *et al.*, 1997; Vogt *et al.*, 1997; Wallon *et al.*, 1997).

One of the most striking features of the *P. aerophilum* HisF structure is the large number of salt bridges on the protein surface, 21 in total for a subunit of 253 residues. Some of these are in extended clusters, one cluster for example involves Arg6, Glu49, Lys102, Glu170 and Asp222 in an extended network. It is also noteworthy that both the N-terminal and C-terminal regions of the polypeptide chain are tied down by salt bridges, involving Arg6 near the N-terminus (which participates in three salt bridges), and Glu246, Arg247, Glu250 and Arg252 near the C-terminus. These interactions may help to prevent unfolding from the chain termini. Of other HisF proteins, only that from *T. maritima* (an organism with an optimum growth temperature of 353 K) is of known structure. Sequence comparisons indicate, however, that many of the salt bridges found in *P. aerophilum* HisF cannot be present in the other HisF proteins listed in Fig. 1, because of residue substitutions; 18 out of 21 are likely to be conserved in HisF from *Aquifex aeolicus* (another hyperthermophilic organism with an optimum growth temperature of 368 K), but no more than 13 out of 21 seem possible in any of the other proteins listed. Of course other interactions may take their place, but from this limited analysis it seems that networks of salt bridges may make a major contribution to thermostability in *P. aerophilum* HisF and probably also in *A. aeoliticus* HisF. A similar phenomenon is seen for the glutamate dehydrogenase from another hyperthermophilic organism, *Pyrococcus furiosus* (Yip *et al.*, 1995).

We gratefully acknowledge Clyde Smith and Peter Haebel for help with data collection, the staff of the Stanford Synchrotron Radiation Laboratory for access to synchrotron data-collection facilities and members of the *P. aerophilum* structural genomics consortium for helpful and stimulating interactions. This work was supported by the Marsden Fund of New Zealand and the Royal Society (London) through a travelling fellowship to MJB.

References

Alifano, P., Fani, R., Lio, P., Lazcano, A., Bazzicalupo, M., Carlomagno, M. S. & Bruni, C. B. (1996). *Microbiol. Rev.* **60**, 44–69.
Ames, B. N. & Garry, B. (1959). *Proc. Natl Acad. Sci. USA*, **45**, 1453–1461.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
Cole, S. T. *et al*. (1998). *Nature (London)*, **393**, 537–544.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Corpet, F. (1988). *Nucleic Acids Res.* **16**, 10881–10890.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 505–524.
Fani, R., Lio, P., Chiarelli, I. & Bazzicalupo, M. (1994). *J. Mol. Evol.* **38**, 489–495.
Green, J. M. & Nichols, B. P. (1991). *J. Biol. Chem.* **266**, 12971–12975.
Höcker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A. & Sterner, R. (2001). *Nature Struct. Biol.* **8**, 32–36.
Jacob, F. & Monod, J. (1961). *J. Mol. Biol.* **3**, 318–356.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Klem, T. J. & Davisson, V. J. (1993). *Biochemistry*, **32**, 5177–5186.
Knapp, S., de Vos, W. M., Rice, D. & Ladenstein, R. (1997). *J. Mol. Biol.* **267**, 916–932.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. & Wilmanns, M. (2000). *Science*, **289**, 1546–1550.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rieder, G., Merrick, M. J., Castorph, H. & Kleiner, D. (1994). *J. Biol. Chem.* **269**, 14386–14390.
Thoma, R., Obmolova, G., Lang, D. A., Schwander, M., Jeno, P., Sterner, R. & Wilmanns, M. (1999). *FEBS. Lett.* **454**, 1–6.
Thoma, R., Schwander, M., Liebl, W., Kirschner, K. & Sterner, R. (1998). *Extremophiles*, **2**, 379–389.
Vogt, G., Woell, S. & Argos, P. (1997). *J. Mol. Biol.* **269**, 631–643.
Volki, P., Markiewicz, P., Baikalov, C., Fitz-Gibbon, S., Stetter, K. O. & Miller, J. H. (1996). *Nucleic Acids Res.* **24**, 4373–4378.
Wallon, G., Kryger, G., Lovett, S. T., Oshima, T., Ringe, D. & Petsko, G. A. (1997). *J. Mol. Biol.* **266**, 1016–1031.
Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. & Rice, D. W. (1995). *Structure*, **3**, 1147–1158.