Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The 2.35 Å structure of the TenA homolog from *Pyrococcus furiosus* supports an enzymatic function in thiamine metabolism

Jordi Benach,[a] William C. Edstrom,[a] Insun Lee,[a] Kalyan Das,[b] Bonnie Cooper,[b] Rong Xiao,[b] Jinfeng Liu,[c] Burkhard Rost,[c] Thomas B. Acton,[b] Gaetano T. Montelione[b] and John F. Hunt[a]*

[a]Department of Biological Sciences and Northeast Structural Genomics Consortium, 702A Fairchild Center, MC2434, Columbia University, New York, NY 10027, USA, [b]Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry and Northeast Structural Genomics Consortium, Rutgers University, Piscataway, NJ 08854, USA, and [c]Department of Biochemistry and Molecular Biophysics and Northeast Structural Genomics Consortium, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA

Correspondence e-mail: jfhunt@biology.columbia.edu

© 2005 International Union of Crystallography
Printed in Denmark – all rights reserved

Received 22 November 2004
Accepted 15 February 2005

PDB References: TenA, 1rtw, r1rtwsf

TenA (transcriptional enhancer A) has been proposed to function as a transcriptional regulator based on observed changes in gene-expression patterns when overexpressed in *Bacillus subtilis*. However, studies of the distribution of proteins involved in thiamine biosynthesis in different fully sequenced genomes have suggested that TenA may be an enzyme involved in thiamine biosynthesis, with a function related to that of the ThiC protein. The crystal structure of PF1337, the TenA homolog from *Pyrococcus furiosus*, is presented here. The protomer comprises a bundle of α-helices with a similar tertiary structure and topology to that of human heme oxygenase-1, even though there is no significant sequence homology. A solvent-sequestered cavity lined by phylogenetically conserved residues is found at the core of this bundle in PF1337 and this cavity is observed to contain electron density for 4-amino-5-hydroxymethyl-2-methyl-pyrimidine phosphate, the product of the ThiC enzyme. In contrast, the modestly acidic surface of PF1337 shows minimal levels of sequence conservation and a dearth of the basic residues that are typically involved in DNA binding in transcription factors. Without significant conservation of its surface properties, TenA is unlikely to mediate functionally important protein–protein or protein–DNA interactions. Therefore, the crystal structure of PF1337 supports the hypothesis that TenA homologs have an indirect effect in altering gene-expression patterns and function instead as enzymes involved in thiamine metabolism.

1. Introduction

The *tenA* locus was initially identified in a screen (Pang *et al.*, 1991) for *Bacillus subtilis* genes whose expression produces strong pleiotropic increases in the yield of secreted enzymes (Priest, 1977). Because this effect was shown to be associated with increased transcription of the genes encoding these extracellular enzymes, the locus was named 'transcriptional enhancer A' or *tenA*, and it was hypothesized that the TenA protein product might function as a transcriptional regulatory factor. If this hypothesis is correct, TenA would provide a very rare example of a transcription factor that functions in all three kingdoms of life, as homologous domains are found in proteins in eubacteria, archaebacteria and eukaryotes (Pang *et al.*, 1991).

However, studies of gene distribution in fully sequenced genomes have suggested that TenA may instead function as an enzyme involved in thiamine (vitamin B_1) biosynthesis. This cofactor is an essential nutrient for mammals, in which a nutritional deficiency causes the disease beriberi, but it is synthesized *de novo* by many microorganisms including *Escherichia coli*, *Salmonella typhimurium*, *Bacillus subtilis* and

Saccharomyces cerevisiae. Thiamine pyrophosphate (THI-PP), the active form of this cofactor, has well characterized enzymatic functions in carbohydrate metabolism and more generally in stabilizing acyl carbanions in metabolic transformations (Schowen, 1998).

While more than 30 years of genetic, enzymological (Begley *et al.*, 1999; Hohmann & Meacock, 1998) and structural (Chiu *et al.*, 1999; Settembre *et al.*, 2003) research on thiamine biosynthesis has revealed a plethora of novel chemistry, some catalytic steps are not clearly understood. The thiamine-biosynthesis pathway is divided into two branches directing the synthesis of the two main building blocks of the molecule, the thiazole and pyrimidine moieties. In *E. coli*, the active form of thiamine or thiamine-pyrophosphate is obtained by phosphorylation of thiamine-phosphate by the kinase ThiL. Thiamine phosphate is assembled by ThiE (thiamine phosphate synthase) using two substrates: 4-methyl-5-(β-hydroxyethyl)thiazole phosphate and 4-amino-5-hydroxymethyl-2-methylpyrimidine-pyrophosphate (HMP-PP). The first of these precursors can be synthesized either through phosphorylation of 4-methyl-5-(β-hydroxyethyl)thiazole by the ThiM kinase or through a complex but well established reaction catalyzed by six enzymes (dxs, ThiF, ThiG, ThiI, ThiH and ThiS) employing glyceraldehyde-3-phosphate, pyruvate, cysteine and ATP as substrates (Begley *et al.*, 1999). The second precursor (HMP-PP) is synthesized through phosphorylation of 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMP-P) by the ThiD kinase. HMP-P itself can be synthesized either through phosphorylation of 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) by ThiD or alternatively through a poorly understood reaction catalyzed by the ThiC enzyme using 5′-phosphoribosyl-5-aminoimidazole (AIR) as a substrate. Although isotopic tracer experiments establish AIR as the substrate for HMP synthesis in bacterial extracts, it has not yet been possible to reconstitute this reaction *in vitro* using purified enzymes (Cheng *et al.*, 2002; Estramareix & David, 1990; Himmeldirk *et al.*, 1998; Begley *et al.*, 1999), indicating that it involves kinetic or mechanistic complexities that have yet to be elucidated.

Analysis of the gene content of a series of fully sequenced microbial genomes revealed that some of the well characterized enzymes in the pathway are missing in some organisms capable of *de novo* thiamine biosynthesis (Morett *et al.*, 2003), suggesting that alternative enzymes may function to mediate some catalytic steps. In this context, it was intriguing that the TenA protein is transcribed as part of an operon involved in thiamine biosynthesis in many organisms, including *B. subtilis*. Moreover, this protein domain is found in a translational fusion in the Thi-4 protein from *Neurospora crassa* (Akiyama & Nakashima, 1996), where the second domain in the covalent fusion is the ThiD kinase. Finally, studies of genomic organization (Rodionov *et al.*, 2002; Morett *et al.*, 2003) revealed that the *tenA* gene occurs in different organisms in a largely mutually exclusive manner with the *thiC* gene (which encodes the ThiC enzyme). Subsequent studies showed that the YCR020c (PET18) TenA homologue from yeast could genetically complement a *thiC* deletion in *E. coli*, supporting the hypothesis that the TenA protein may be an enzyme directly involved in *de novo* thiamine biosynthesis with a function related to that of ThiC, despite the absence of sequence homology between these proteins (Morett *et al.*, 2003).

In this paper, we report the crystal structure of the PF1337 protein from *Pyrococcus furiosus*, which is target PfR34 of the Northeast Structural Genomics Consortium. While the surface of this TenA ortholog lacks the basic regions characteristic of transcription factors that bind DNA, a solvent-sequestered cavity is found in the core of the protomer containing electron density for 4-amino-5-hydroxymethyl-2-methyl-pyrimidine phosphate (HMP-P), an intermediate in the thiamine-biosynthesis pathway that is the product of ThiC. Therefore, the crystal structure of TenA strongly supports an enzymatic function in thiamine biosynthesis and casts doubt upon the hypothesis that it directly functions as a transcription factor. These results demonstrate the utility of protein structure determination in clarifying biological and enzymatic function, which represents a major goal of contemporary genomic biology.

2. Materials and methods

2.1. Protein expression, purification and crystallization

Selenomethionine-labeled full-length PF1337 was expressed and purified with an eight-residue C-terminal affinity tag (LEHHHHHH) using previously described methods (Benach *et al.*, 2003). In brief, a pET21d (Novagen) derivative was used for T7 expression in *E. coli* BL21(DE3) cells containing a rare tRNA expression plasmid, and the protein was purified using nickel-affinity and gel-filtration chromatography prior to concentration to 8 mg ml^{-1} in 10 m*M* Tris, 100 m*M* NaCl and 5 m*M* DTT pH 8.0. Sample purity (>97%) and molecular weight (25.7 kDa) were verified by SDS–PAGE and matrix-assisted laser desorption ionization time-of-flight mass spectroscopy (MALDI–TOF), respectively. Orthorhombic plate-shaped crystals grew at 277 K in 3 d to dimensions of $\sim$50 $\times$ 100 $\times$ 400 µm in 1 + 1 µl hanging-drop vapor-diffusion reactions over a reservoir containing 20% PEG 400, 200 m*M* $CaCl_2$, 1 m*M* DTT, 50 m*M* cacodylate pH 7.5. Crystals were frozen in liquid propane using an additional 10% PEG 400. Owing to the limited commercial availability of HMP(-P) compounds, attempts to co-crystallize with HMP(-P) were not attempted.

2.2. X-ray data collection and structure determination

Multi-wavelength anomalous diffraction data were collected from a single crystal at the selenium edge on a Brandeis B4 detector on beamline X12C at the National Synchrotron Light Source in consecutive 370° sweeps at 0.9790 Å (peak), 0.9794 Å (edge) and 0.9500 Å (remote) using 1°, 30 s oscillations. Data were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). For each data set, the scaling *B* factors of the final frames were within 3 Å^2 of the initial frames, indicating that there was minimal decay

during data collection. *SOLVE* (Terwilliger & Berendzen, 1999) identified 17 of the 20 Se atoms in the tetramer in the asymmetric unit, yielding a map that was used for non-crystallographic symmetry (NCS) averaging, density modification (with 48% solvent content) and automated iterative model building in *AUTO_RESOLVE* (Terwilliger & Berendzen, 1999; Murshudov *et al.*, 1997). This program identified 71% of the backbone and 64% of the side chains in the final model and produced a map that enabled 97% of the structure to be built by hand using *O* (Jones *et al.*, 1991). Refinement in *CNS* (Brünger *et al.*, 1998) was conducted as previously described (Benach *et al.*, 2003) using standard geometric and van der Waals restraints (Engh & Huber, 1991) and was monitored by a randomly selected $R_{\rm free}$ set containing 5% of the reflections. *B* factors were refined using standard vicinal restraints (1.5–2.0 Å^2 for main-chain atoms and 2.0–2.5 Å^2 for side-chain atoms). Strong NCS restraints (1050 kJ Å^{-2}, σ_B = 1.5) were initially maintained throughout the model but ultimately were released competely to produce a considerable improvement in $R_{\rm free}$. The final model contained residues 1–209 in molecule *B* (*i.e.* the entire protein excluding the tag), residues 1–208 in molecule *C* and residues 1–206 in molecules *A* and *D*. Water-molecule sites were selected automatically using *CNS* and checked for consistency with $2F_{\rm o} - F_{\rm c}$ electron-density and hydrogen-bonding criteria.

2.3. HPLC analysis of protein-bound small molecules

50 µg protein from the crystallization stock was lyophilized after dialysis against 1 m*M* imidazole pH 7.0 and then dissolved in 90% formic acid prior to injection onto a Vydac C18 column (5 µm particles, 1.0 mm internal diameter, 150 mm length) running at 1 ml min^{-1} in 1% acetonitrile (AcN), 0.1% formic acid. A linear gradient to 80% AcN, 0.1% formic acid was run after a 2 min isocratic elution. A compound with a pyrimidine-like absorbance around 260 nm eluted just before the start of the gradient in the protein sample but not in a matched sample prepared from an equal volume of dialysis buffer (data not shown).

3. Results and discussion

3.1. The crystal structure of PF1337

The PF1337 protein from *P. furiosus* shows 23% identity and 48% similarity to the TenA protein from *B. subtilis*. Some kind of orthologous relationship between these proteins is supported by this sequence homology combined with the fact that the *pf1337* gene is located in a thiamine-biosynthesis operon in *P. furiosus* and the fact that the genome of this organism does not encode any other proteins with significant homology to TenA. The crystal structure of PF1337 was solved using multiwavelength anomalous diffraction (MAD) from selenomethionine-labeled protein and refined to $R_{\rm work}$ = 0.24 and $R_{\rm free}$ = 0.28 at 2.35 Å resolution (Table 1). The secondary structure of the protomer is entirely α-helical (Figs. 1*a* and 1*b*). Its core comprises a six-helix bundle surrounding a central cavity that contains strong electron density in one protomer

Table 1
P. furiosus PF1337 data-collection and refinement statistics.

Standard definitions were used for all parameters (Drenth, 1994). Values in parentheses are for the limiting resolution shell (2.43–2.35 Å). The data-reduction statistics are from *SCALEPACK* (Otwinowski & Minor, 1997) and the refinement statistics from *CNS* (Brünger *et al.*, 1998). HMP-P is an abbreviation for 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate.

Crystal parameters	
Space group	$P2_12_12$
Unit-cell parameters (Å) at 100 K	a = 101.8, b = 123.6, c = 77.5
Data quality	
Resolution range (Å)	20.0–2.35
No. of measured reflections	239958 (22349)
No. of unique reflections	73637 (6861)
$R_{\rm sym}$ [$I \geq -3\sigma(I)$]	0.07 (0.33)
Mean $I/\sigma(I)$ [$I \geq \sigma(I)$ after merging]	18.7 (6.6)
Mean redundancy	3.3 (3.3)
Completeness (%)	
Measured reflections	93.5 (89.7)
$I \geq 2\sigma(I)$ (%)	74.8 (51.8)
Refinement residuals [$F \geq \sigma(F)$]	
$R_{\rm free}$	0.28 (0.31)
$R_{\rm work}$	0.24 (0.27)
Model quality	
R.m.s.d. bond lengths (Å)	0.007
R.m.s.d. bond angles (°)	1.1
Average *B* factors (Å^2)	
All	32.6
Main chain	30.3
Side chain	34.4
Average *B* factor for waters (Å^2)	34.2
R.m.s.d. *B* factors for nonbonded contacts (Å^2)	
Main chain	4.9
Side chain	9.2
Ramachandran plot (%)	
Core	96.1
Additional allowed	3.9
Model contents	
No. of protein atoms	7089
No. of HMP-P atoms	14
No. of phosphate atoms	15
No. of water molecules	234
Protein residues	206 + 209 + 208 + 206
HMP-P molecules	1
Phosphate molecules	3

consistent with the presence of HMP-P. (See below for a discussion of ligand identification.) There are three short α-helices at the N-terminus of the protomer that pack onto one surface of the helical bundle forming the core. Furthermore, one of the α-helical segments forming the core contains a short disruption that breaks it into two separate α-helices according to standard secondary-structure assignment programs (α8 and α9 in Fig. 1*a*), so that the structure contains a total of ten α-helices. Three of the five connections between the helices forming the six-helix bundle are non-local (*i.e.* between helices that are not at adjacent positions in the bundle), giving the bundle a relatively complex topology (Fig. 1*b*) involving both parallel and antiparallel α-helix packing (Fig. 1*a*).

The asymmetric unit of the crystal contains a tetramer of PF1337 (Fig. 1*c*) with 222 symmetry (*i.e.* with three orthogonal twofold axes) that buries 3528 Å^2 of solvent-accessible surface area per protomer as calculated by *CNS* (Brünger *et al.*, 1998). This oligomer is likely to be physiological based on the fact that gel filtration and static light-scattering indicate PF1337 to

be tetrameric in solution (see supplementary material).[1] Furthermore, a similar tetramer is found in the structure of PH1161, the TenA ortholog from *P. horikoshii* (PDB code 1udd; Itou *et al.*, 2004). One of the twofold axes in the tetramer involves a tight dimer interaction in which 1600 Å^2 per protomer is buried in a reciprocal packing interface between helices α4 and α5 in subunits *A* and *B* (or *C* and *D*). The second twofold axis buries 1159 Å^2 per protomer in a reciprocal packing interaction between helix α6 in subunits *A* and *D* (or *B* and *C*), while the third buries 963 Å^2 per protomer in a reciprocal packing interaction between the loops connecting helices α4–5, α6–7 and α9–10 in subunits *A* and *C* (or *B* and *D*). Superimposing the individual monomers in the tetramer gives r.m.s.d.s for C$^\alpha$ atoms of 0.21–0.30 Å, with the largest differences observed in the loops between helices α3–4, α5–6 and α7–8. The 222 symmetry is most exact between the *AB*/*DC* dimers (179.8°) and least exact between the *AC*/*BD* dimers (178.4°).

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: SX5026). Services for accessing these data are described at the back of the journal.

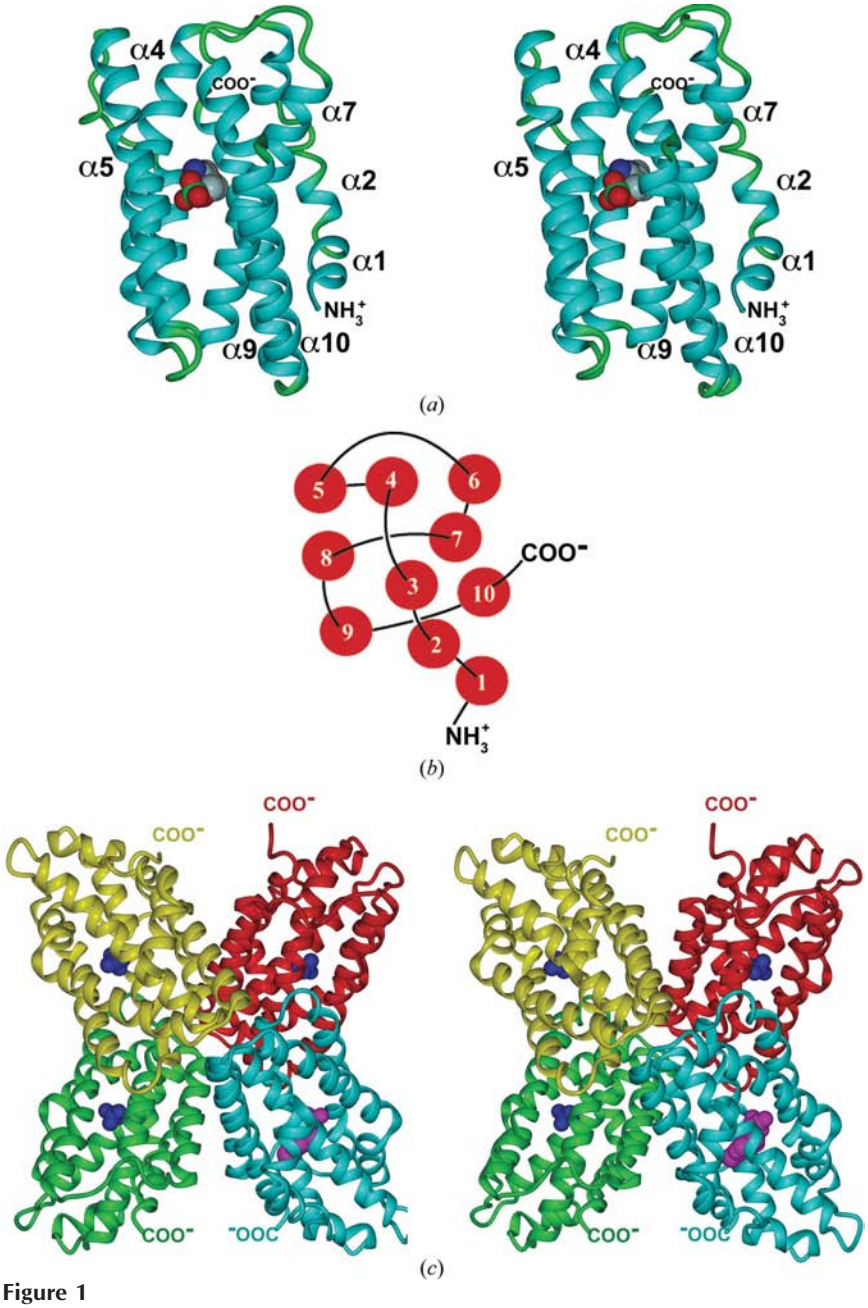


Figure 1
Crystal structure of PF1337. (*a*) Stereo ribbon diagram (Carson, 1987) of PF1337 with a CPK representation of HMP-P (oxygen, red; nitrogen, blue; carbon, gray; phosphorus, green). (*b*) Topology diagram of PF1337. Red circles represent α-helices, oriented perpendicular to the plane of the page, with connections above the plane penetrating into the center of the circle and those below stopping at the boundary. (*c*) Stereo ribbon diagram of the PF1337 tetramer with CPK representations of HMP-P (magenta, subunit *A*) and inorganic phosphate (blue, subunits *B*–*D*). The different subunits are colored in cyan (*A*), red (*B*), yellow (*C*) and green (*D*).

3.2. Similarity of the PF1337 fold to known structures

A search for structures similar to PF1337 was conducted using the program *DALI*, which identified numerous domains with similar tertiary structure. Most of these bind heme or an iron ion and function in mediating redox reactions. The closest similarity to a protein that is not likely to be orthologous to TenA is to human heme oxygenase-1 (PDB code 1n45; Schuller *et al.*, 1999), which gives a *Z* score of 16.4 and an r.m.s.d. of 3.1 Å for the alignment of the C$^\alpha$ atoms in 189 residues sharing 14% sequence identity (Fig. 2*a*). This protein has identical topology to PF1337 and catalyzes the NADPH-dependent oxidation of a heme molecule that is sandwiched between its amino-terminal α-helix and the helical bundle forming its core (Fig. 2*a*). Strong structural similarity is also observed to the R2 subunit of ribonucleotide reductase (PDB code 1r2f; Logan *et al.*, 1996; *Z* score of 11.2 for alignment of 174 C$^\alpha$ atoms with 3.1 Å r.m.s.d. and 12% sequence identity) and to methane monooxygenase hydroxylase (PDB code 1mhy; Elango *et al.*, 1997; *Z* score of 10.5 for alignment of 171 C$^\alpha$ atoms with 3.1 Å r.m.s.d. and 12% sequence identity). While these proteins have a different topology to PF1337, they both bind redox-active di-iron ions at a site coincident with the cavity containing HMP-P in PF1337 (*e.g.* as shown in Fig. 2*b*). While we find no evidence for the presence of iron ion or an iron-binding motif in PF1337, the shared architecture could reflect the ability of this protein fold to provide a completely sequestered active site to protect solvent-sensitive enzymatic intermediates.

3.3. Sequence-conservation patterns suggests that PF1337 is an enzyme

The surface of PF1337 is mildly acidic (second row in Fig. 3*a*), while that of DNA-binding transcription factors is generally basic (Kim *et al.*, 1993), and the relatively sparse basic residues on the surface of PF1337 are not phylogenetically conserved (second and third rows of Fig. 3*a* and Fig. 4). Moreover, PF1337 shows minimal levels of phylogenetic conservation of any of its solvent-exposed surface residues (third row of Fig. 3*a*). The highest sequence conservation on the surface of the protomer occurs in the intersubunit interfaces of the tetramer, but even the most strongly conserved of these sites (Ser114) has an identical residue in only 65% of the homologs. In the non-interface region, there is one partially exposed residue that is very strongly conserved (His20), but the side chain of this residue forms a partially buried salt bridge. The very low level of sequence conservation among the solvent-exposed residues in PF1337 is contrary to the pattern observed in proteins that bind macromolecular partners such as protein or DNA (*e.g.* as shown in Figs. 3*b* and 3*c*; Kim *et al.*, 1993; Bahadur *et al.*, 2004; Halperin *et al.*, 2004, Fernandez-Recio *et al.*, 2004). In contrast, residues in the core of PF1337 are very strongly conserved (fourth row of Fig. 3*a*). In particular, many of the acidic and aromatic residues lining the ~330 Å^3 cavity that putatively binds HMP-P are nearly invariant in all TenA homologs (fourth row of Fig. 3*a* and Figs. 4 and 5*b*). Therefore, the pattern of sequence conservation in PF1337 is consistent with function as an enzyme with specificity for HMP-P and contrary to expectations for a protein that functions as a transcriptional regulator mediating either protein–protein or protein–DNA interactions.

While many transcription factors are known to bind small-molecule cofactors, these proteins show a high level of phylogenetic conservation in their surface residues just like other proteins involved in mediating protein–protein or protein–DNA interactions (Bahadur *et al.*, 2004; Halperin *et al.*, 2004; Fernandez-Recio *et al.*, 2004). Two specific examples of the surface-sequence conservation in ligand-dependent transcription factors are shown in Figs. 3(*b*) and 3(*c*). The NAD(P)(H)-dependent NmrA protein from *Aspergillus nidulans* regulates transcription *via* interactions with the AreE protein (Stammers *et al.*, 2001) and the protein–protein interaction interfaces of NmrA show strong phylogenetic conservation (Fig. 3*b*). The cAMP-dependent CAP protein binds both a specific DNA sequence and the C-terminal domain of RNA polymerase (Schultz *et al.*, 1991; Benoff *et al.*, 2002) and its surfaces forming either the protein- or DNA-binding sites exhibit similarly strong phylogenetic conservation (Fig. 3*c*). No comparable pattern of sequence conservation is observed anywhere on the surface of PF1337 when analyzed using the same methods (third row in Fig. 3*a*), and this inconsistency with the observed properties of other transcription factors makes it unlikely that TenA functions in this capacity.

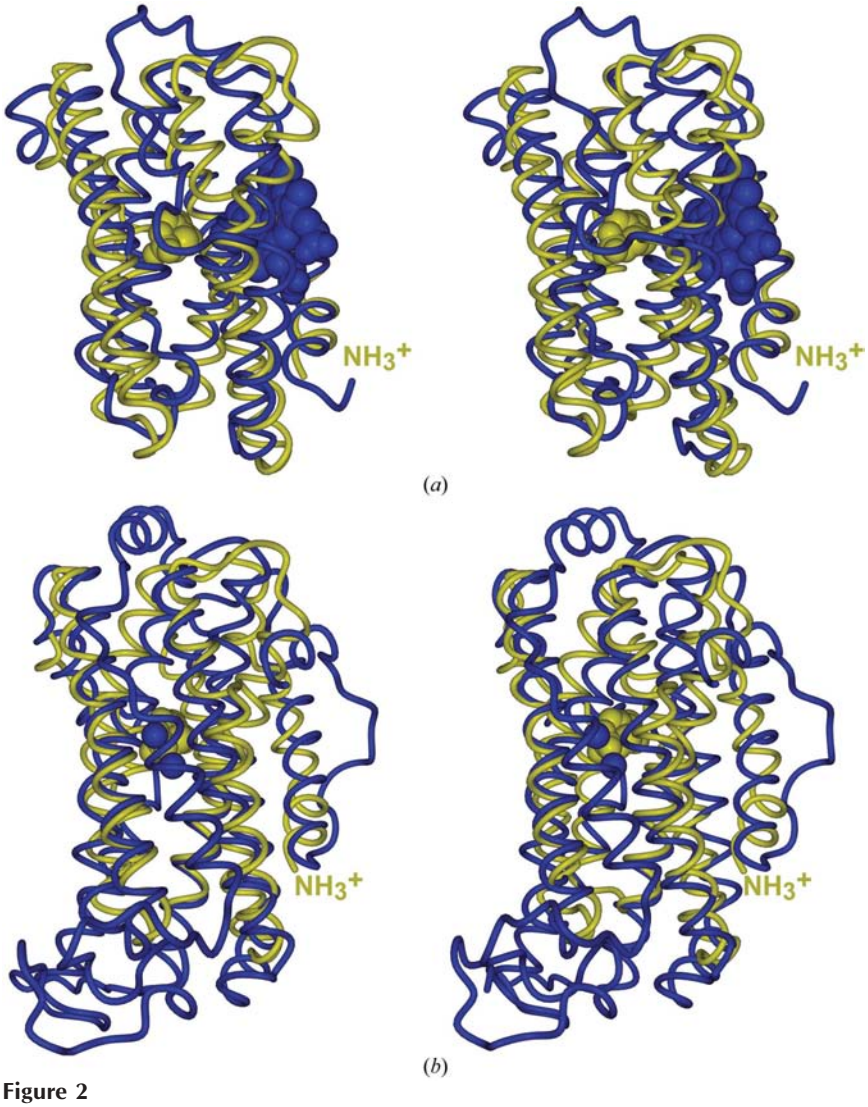


Figure 2
Comparison to similar structures. Stereo diagrams showing superposition of PF1337 (yellow) and structural homologs (blue). (*a*) shows human heme oxygenase (PDB code 1n45, with heme ligand), while (*b*) shows *E. coli* ribonucleoside-diphosphate reductase 1 β-chain (PDB code 1xik, with di-iron ligands). Bound ligands are shown in CPK representations colored like the protein backbone (with HMP-P in PF1337). Figures were produced with *DINO* (http://www.dino3d.org) and rendered with *POVRAY* (http://www.povray.org).

3.4. Modeling of an HMP-P molecule in the putative active site

Following completion of the refinement of the PF1337 protein structure, strong residual electron density was visible in the cavity lined by phylogenetically conserved residues at the core of the protomer (Figs. 5*a*, 5*b* and 5*c*). This density was consistent with the presence of a molecule

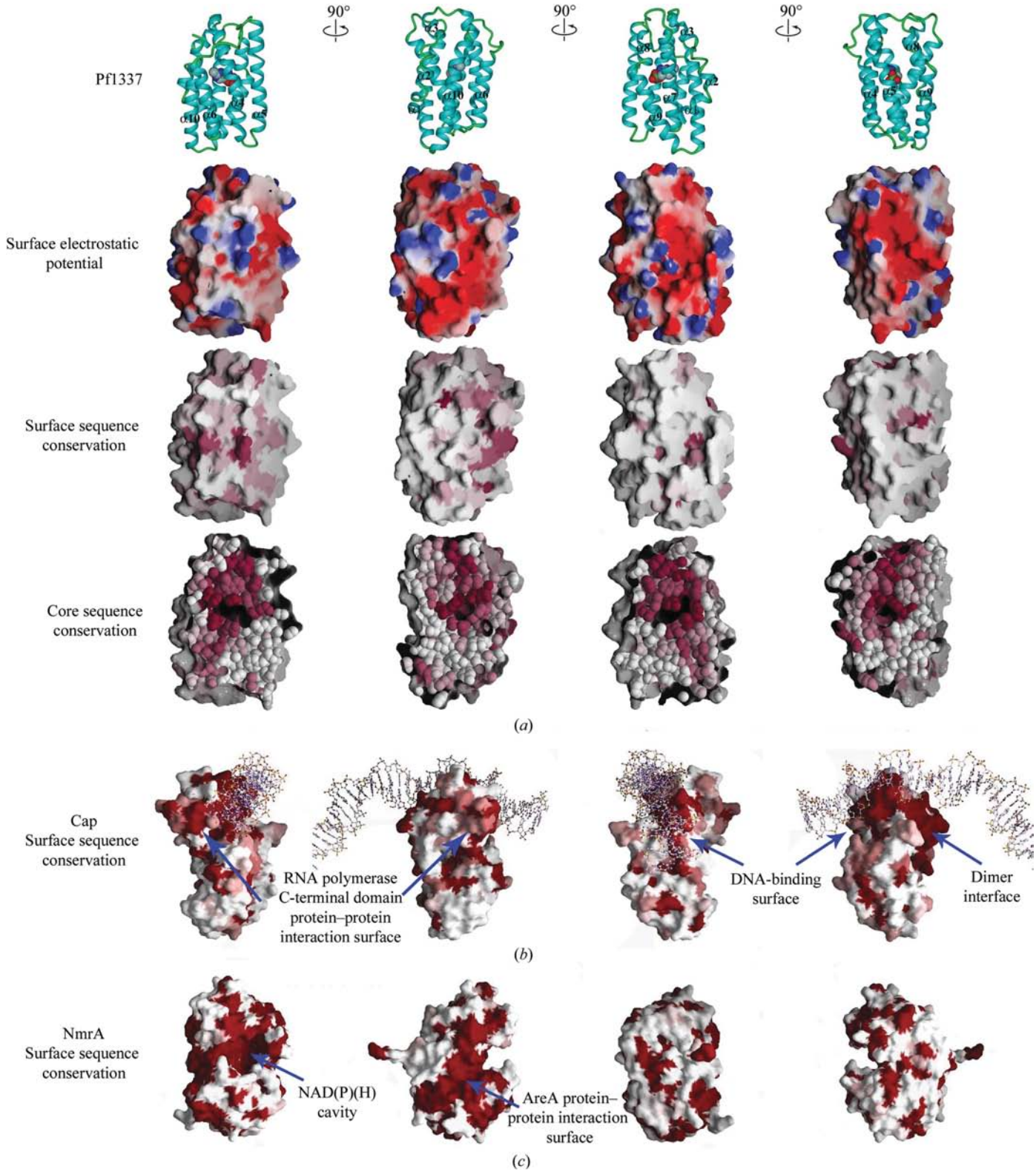


Figure 3
Four views of the surface properties of PF1337, *E. coli* CAP and *A. nidulans* NmrA. (*a*) Four views of the PF1337 protomer showing the accessible surface (Nicholls *et al.*, 1991) colored by electrostatic potential (second row, with red or blue indicating potentials at 100 m*M* salt of $\leq -5kT$ or $\geq +5kT$, respectively) or sequence conservation [third row, with maroon indicating 100% conservation and white indicating no conservation among the sequences in the COG (cluster of orthologous groups; Tatusov *et al.*, 1997) according to the *CONSURF* algorithm (Glaser *et al.*, 2003)]. The fourth row presents cut-away views of a CPK model of the core of the protomer showing sequence conservation (analyzed and colored in the same manner as for the third row). For reference, the first row shows ribbon diagrams of the protomer colored as in Fig. 1(*a*). (*b*) *CONSURF* analysis of *A. nidulans* NmrA (PDB code 1k6x) showing the regions that bind NAD(P)(H) and AreA. (*c*) *CONSURF* analysis of catabolite gene activator protein (CAP) from *E. coli* complexed with DNA (PDB code 1cgp).

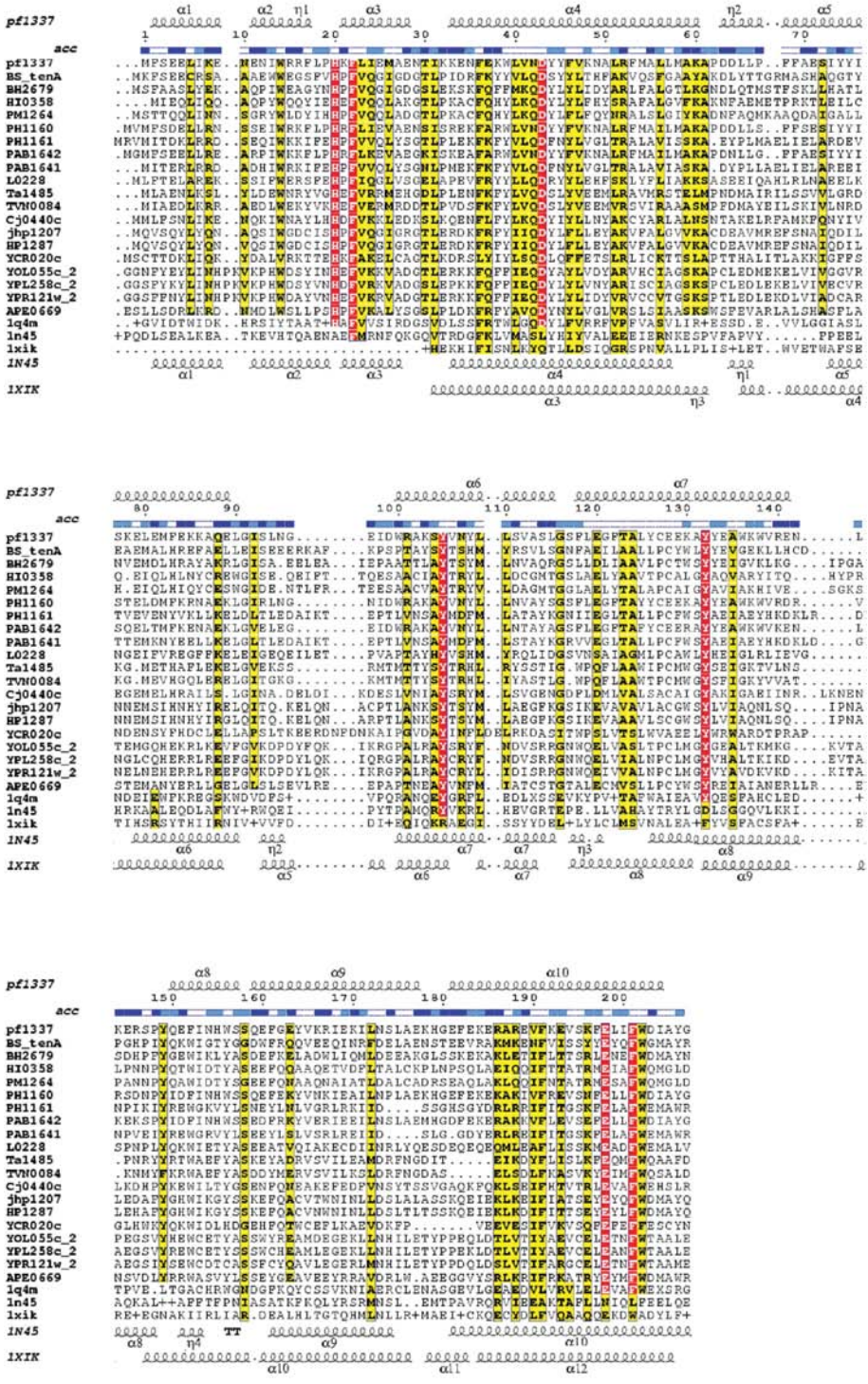
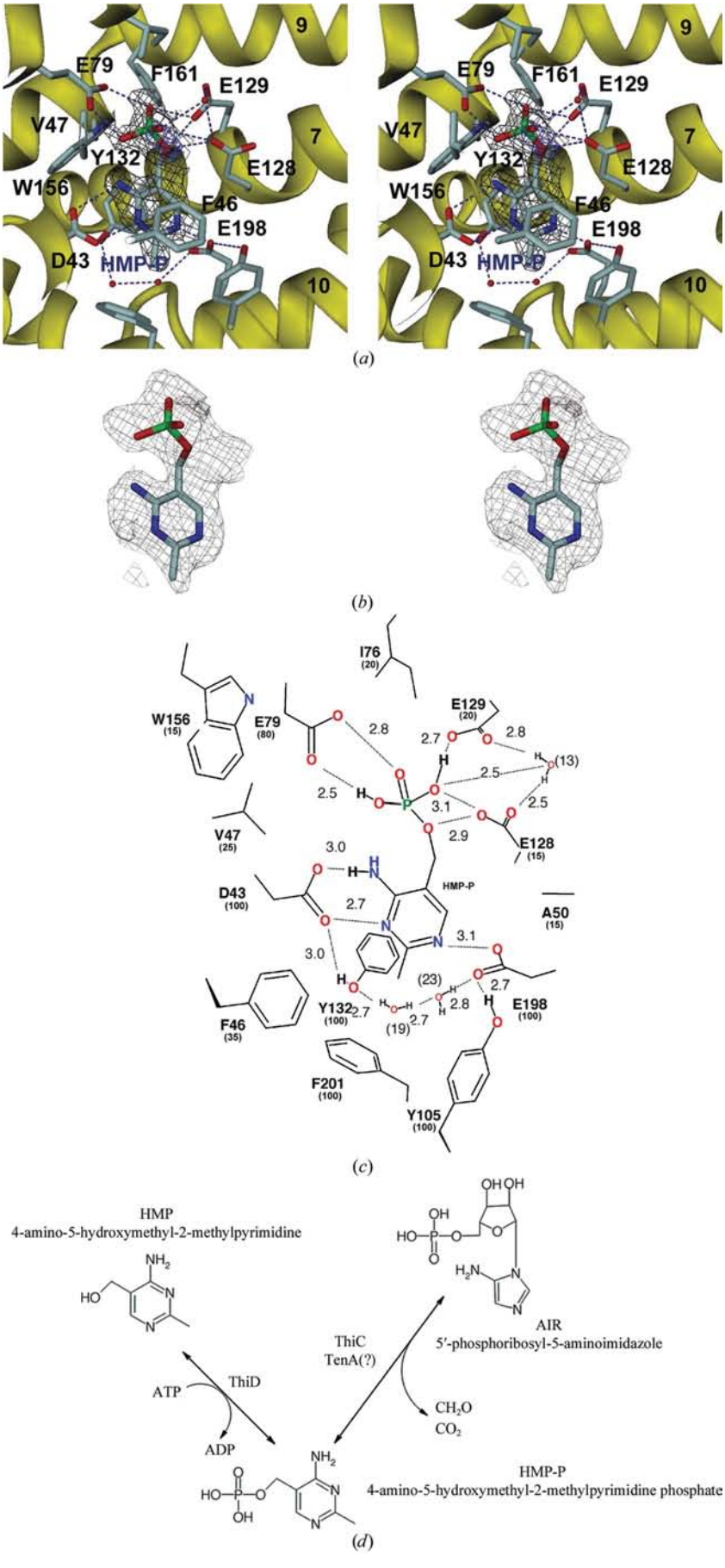


Figure 4
Phylogenetic conservation in the PF1337 sequence and structure. Sequence-structure alignment with members of the same cluster of orthologous genes (COG; Tatusov *et al.*, 1997) made using *ESPRIPT* (Gouet *et al.*, 1999). Strictly conserved and conservatively substituted residues are colored with a red and yellow background, respectively [according to *ESPRIPT* using a residue-similarity Risler matrix; Risler *et al.*, 1988)]. The secondary structure of PF1337 is shown at the top, with coils representing α-helices, and the colored bar indicates relative solvent-accessibility, with blue indicating accessible, white buried and cyan intermediate. Sequences are labeled with SWISS-PROT ids or standard genomic open reading frame numbers. Three structure-based sequence alignments are shown at the bottom, labeled by PDB code. 1q4m is the *A. thaliana* seed-maturation protein, a TenA homolog, while 1n45 is human heme oxygenase-1 (Schuller *et al.*, 1999), and 1xik is the R2 subunit of ribonucleotide reductase from *E. coli* (Logan *et al.*, 1996). The secondary structures of 1n45 and 1xik are shown at the bottom.

of HMP-P in subunit *A* and inorganic phosphate molecules in the other subunits. We excluded the presence of an Fe iron or a cacodylate molecule based on the absence of peaks in anomalous difference Fourier maps (data not shown). The correct identification and high occupancy of the HMP-P ligand is supported by several factors, including low residual electron density and good molecular geometry in the ligand after refinement and the proximity of its *B* factors to those of the adjacent protein side chains (with a mean of 37 Å^2 for HMP-P compared with $\sim$20 Å^2 for neighboring atoms). Furthermore, acid extracts of the protein in the crystallization stock confirmed the presence of a pyrimidine (see §2.2).

The HMP-P makes numerous contacts to strongly conserved protein residues (Figs. 5*a* and 5*c*). The pyrimidine ring is sandwiched between the aromatic side chains of Tyr132 and Phe46, and the methyl group at position 2 makes van der Waals contacts to the side chain of Phe201. Moreover, there are numerous protein–ligand hydrogen bonds, including ones with heteroatom distances of $\leq$3.1 Å to all the N atoms of HMP-P (Fig. 5*c*). Interestingly, the hydrogen bonds between the side chain of Asp43 and the N4 and amino groups of the ligand mimic those observed in HMP-P kinase (Cheng *et al.*, 2002). Glu79, Glu128 and Glu129 all make hydrogen bonds to the phosphate in the ligand. The fact that the most highly conserved amino acids in this putative active site are clustered around the pyrimidine moiety in HMP-P (Fig. 5*c*) suggests that PF1337 and other TenA homologs may catalyze the formation or chemical modification of the pyrimidine ring, as previously proposed for ThiC (Cheng *et al.*, 2002; Estramareix & David, 1990; Himmeldirk *et al.*, 1998; Begley *et al.*, 1999). Interestingly, the residues ligating the phosphate moiety of HMP-P are substantially less conserved, with the exception of Glu79 (which occurs in 80% of the likely orthologs; Figs. 4 and 5*c*).

This observation suggests that different TenA homologues might have specificity for different compounds



sharing a common pyrimidine moiety, thereby raising the possibility of functional heterogeneity among TenA homologues. Most importantly, the many complementary interactions observed at the putative active site in PF1337 suggest that this protein has high specificity for HMP-P, consistent with its co-purification with the protein. The very slow rate of product release implied by the co-purification of HMP-P with TenA suggests that the enzyme will turn over very slowly (at least in the absence of additional molecular interactions), and slow enzyme turnover could explain why it has been difficult to observe enzymatic activity from the purified protein (Costello, 1996; Begley *et al.*, 1999).

The co-purification of HMP-P with the PF1337 protein is probably promoted by the fact that its binding site is completely solvent-sequestered (third row in Fig. 3*a* and Fig. 6*b*), meaning that substantial fluctuations in protein structure would be required for ligand release. Interestingly, the crystal structure of a TenA homolog from *P. horikoshii* (PDB code 1udd) does not show any electron density for HMP-P in its core, despite showing a 1.7 Å r.m.s.d. for the superposition of 202 of 206 C^{α} atoms in PF1337 in residues sharing 27% sequence identity. The most significant difference is a 5 Å lateral movement of helix α5 relative to helix α9 (Fig. 6*a*), which results in solvent exposure of the HMP-P-binding cavity (Fig. 6*b*). This movement could allow substrate binding and product release from the putative active site in PF1337 and other TenA homologs.

Figure 5
The putative active site and function of PF1337. (*a*) Stereo pair showing the environment of the crystallographically observed HMP-P molecule bound to PF1337. Residues within a 4 Å radius of HMP-P atoms are shown. The gray electron density is from an unbiased $2F_o - F_c$ map calculated prior to the inclusion of HMP-P in the refinement and contoured at 1σ. (*b*) Stereo pair of the electron density of an unbiased $F_o - F_c$ map contoured at 1.8σ. (*c*) Schematic plot of the molecular interactions of the bound HMP-P. Distances are given in angstroms, and the values in parentheses represent the percentage identity among the TenA sequences in the COG in the case of protein residues or *B* factors in Å^2 in the case of water molecules. The phosphate moiety of HMP-P is assumed to be in a protonated state, presumably reflecting a perturbation in its pK_a caused by the chemical environment in the active site. (*d*) Reaction scheme for the biosynthesis of HMP-P from either HMP or AIR. TenA is hypothesized to catalyze a similar reaction to ThiC, consistent with the ability of the *tenA* gene to genetically complement a *thiC* deletion (Morett *et al.*, 2003).

3.5. Implications for the biochemical function of TenA homologs

The identification of an HMP-P molecule in the phylogenetically conserved cavity at the core of PF1337, combined with the minimal level of sequence conservation on its surface, strongly supports the hypothesis that TenA homologs play an enzymatic role in thiamine metabolism. Phosphorylation of HMP-P yields one of the two basic building blocks of thiamine. Based on the current understanding of microbial thiamine-biosynthesis pathways, there are two different routes to the synthesis of HMP-P, one involving the phosphorylation of 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) by the ThiD protein and the other involving a hypothesized multistep rearrangement of AIR by the ThiC enzyme (Fig. 5*d*). The binding of HMP-P to the PF1337 homolog whose structure is reported in this paper combined with the previously demonstrated ability of another TenA homolog to complement a *ThiC* mutation *in vivo* suggests that some homologs may catalyze the same chemical transformation(s) as ThiC to produce HMP-P.

However, the exact chemical transformation performed by ThiC has not been established because it has not been possible to demonstrate its activity in purified enzyme preparations (Costello, 1996). The hypothesis that ThiC converts AIR to HMP-P is based on isotopic tracer experiments performed using crude cellular extracts, which show that the 4′, 5′ and 2′ ribose C atoms of AIR are converted into the C5, C7 and C8 atoms of the pyrimidine ring in HMP-P (Cheng *et al.*, 2002; Estramareix & David, 1990; Himmeldirk *et al.*, 1998; Begley *et al.*, 1999). A complex 11-step reaction mechanism has been proposed to explain this remarkable transformation on ThiC. Because all proposed steps involve acid–base reactions, such a complex chemical transformation could conceivably be performed in the active site of PF1337 and other TenA homologs. While the volume of the active-site cavity in the PF1337 structure seems to be too small to accommodate an AIR molecule, the volume of the cavity could dynamically expand in the course of the enzymatic reaction coupled with specific protein conformational changes in the course of the enzymatic reaction cycle. Moreover, the complete solvent sequestration of the active site in PF1337 could be explained by the need to prevent release of the reactive chemical fragments that are proposed to be formed and then recombined in the course of the proposed 11-step enzyme-bound transformation.

On the other hand, in the absence of a direct demonstration of activity by purified enzymes, the exact chemical transformation performed by ThiC or TenA may be different from the hypothesized conversion of AIR to HMP-P. Therefore, HMP-P could conceivably be a substrate for TenA or even an inhibitor. Moreover, the variable stringency of sequence conservation at the pyrimidine-binding and phosphate-binding regions at the active site in PF1337 raises the possibility of functional diversity among TenA homologues. In conclusion, a variety of questions remain to be answered concerning the chemical transformations catalyzed by PF1337 and other TenA homologs and their coupling to the dynamic processes mediating ligand binding and release.

Note added in proof: Toms *et al.* (2005) have just reported the structure of TenA from *B. subtilis* with HMP bound at its active site, and they also presented data showing that this enzyme has thiaminase II activity, which cleaves thiamine into thiazole and HMP. However, this hydrolytic activity does not

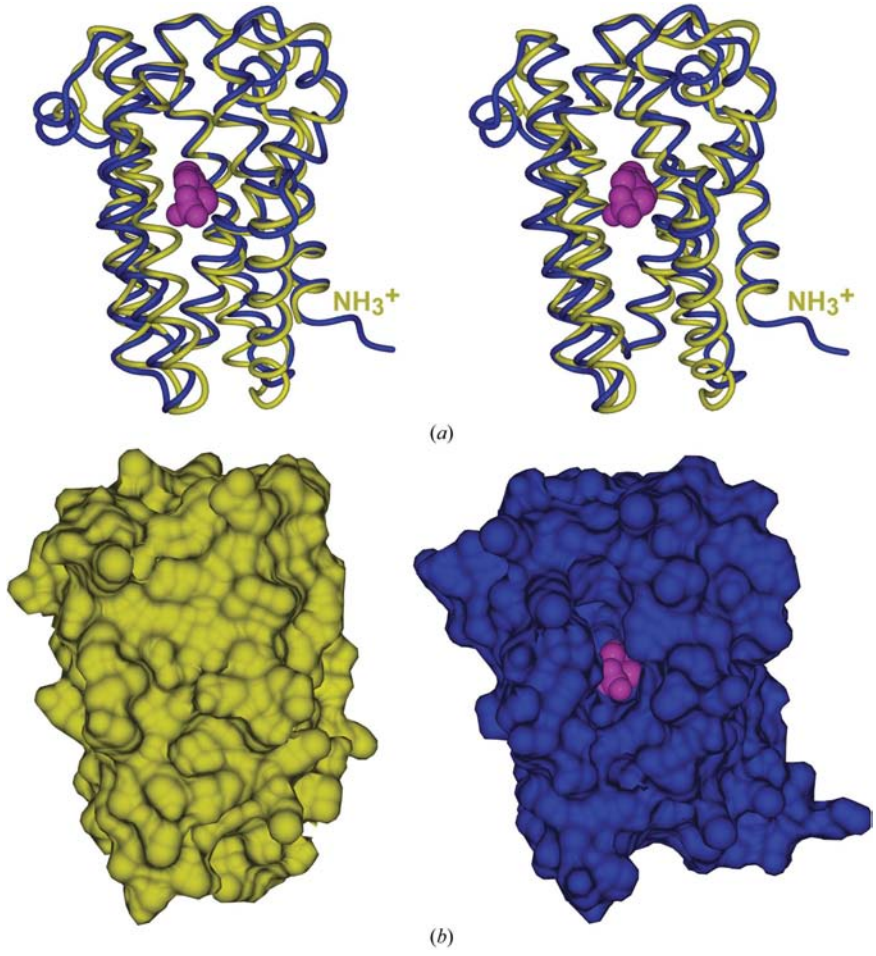


Figure 6
Structural comparison of PF1337 with a *P. horikoshii* homolog. (*a*) Stereo pair (Carson, 1987) showing least-squares alignment of PF1337 (yellow) and PH1161 (blue), with the crystallographically observed HMP-P molecule in PF1337 shown in magenta CPK representation. (*b*) The accessible surfaces of the *P. furiosus* (left) and *P. horikoshii* (right) homologs (in the same orientation as Fig. 6*a*), with HMP-P shown in magenta CPK representation. HMP-P was positioned in the putative active site of the *P. horikoshii* protein based on its alignment with subunit *A* of PF1337. The failure to observe the bound HMP-P molecule in PF1337 demonstrates its sequestration from solvent contact.

explain either the ability of the YCR020c TenA homologue from yeast to complement a thiC deletion (Morett, 2003) or the co-purification of HMP-P with PF1337 as reported in this paper. The active site of PF1337 is too small to accommodate both thiamine and phosphate simultaneously, so it is unlikely to cleave thiamine using inorganic phosphate. Furthermore, YCR020c lacks the Cys residue proposed by Toms *et al.* to initiate the cleavage of thiamine *via* nucleophilic attack on the pyrimidine ring, and the equivalent residue in archaeal TenA homologues including PF1337 (Cys127) is shifted by one position in the protein sequence. This shift causes the S atom of the proposed catalytic residue to be located 9 Å away from the target atom on the pyrimidine ring of HMP-P in the PF1337 structure, but the pyrimidine ring makes otherwise equivalent molecular interactions in the PF1337 complex with HMP-P and the *B. subtilis* TenA complex with HMP. These observations suggest that YCR020c and the archaeal TenA homologues may not have the thiaminase activity observed in the *B. subtilis* enzyme. Most strikingly, the phosphate-ligating residues in PF1337 (Fig. 5*c*) are conserved in all of the archaeal TenA homologues that lack the putative catalytic cysteine at the position observed in *B. subtilis* TenA, strongly reinforcing the possibility of functional diversity among TenA homologues.

This work was supported by a grant from the Protein Structure Initiative of the National Institutes of Health (NIGMS-P50-GM62413). The authors thank Anand Saxena of Brookhaven National Laboratory for support at the synchrotron and Paul LaPlae and Ronald Breslow for stimulating discussions and assistance with HPLC analyses.

References

Akiyama, M. & Nakashima, H. (1996). *Curr. Genet.* **30**, 62–67.

Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. (2004). *J. Mol. Biol.* **336**, 943–955.

Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J. & Xi, J. (1999). *Arch. Microbiol.* **171**, 293–300.

Benach, J., Lee, I., Edstrom, W., Kuzin, A. P., Chiang, Y., Acton, T. B., Montelione, G. T. & Hunt, J. F. (2003). *J. Biol. Chem.* **278**, 19176–19182.

Benoff, B., Yang, H., Lawson, C. L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y. W., Berman, H. M. & Ebright, R. H. (2002). *Science*, **297**, 1562–1566.

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

Carson, M. (1987). *J. Mol. Graph.* **5**, 103–106.

Cheng, G., Bennett, E. M., Begley, T. P. & Ealick, S. E. (2002). *Structure*, **10**, 225–235.

Chiu, H. J., Reddick, J. J., Begley, T. P. & Ealick, S. E. (1999). *Biochemistry*, **38**, 6460–6470.

Costello, C. (1996). PhD Thesis. Cornell University, USA.

Drenth, J. (1994). *Principles of Protein X-ray Crystallography*. New York: Springer–Verlag.

Elango, N., Radhakrishnan, R., Froland, W. A., Wallar, B. J., Earhart, C. A., Lipscomb, J. D. & Ohlendorf, D. H. (1997). *Protein Sci.* **6**, 556–568.

Engh, R. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Estramareix, B. & David, S. (1990). *Biochim. Biophys. Acta*, **1035**, 154–160.

Fernandez-Recio, J., Totrov, M. & Abagyan, R. (2004). *J. Mol. Biol.* **335**, 843–865.

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). *Bioinformatics*, **19**, 163–164.

Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.

Halperin, I., Wolfson, H. & Nussinov, R. (2004). *Structure*, **12**, 1027–1038.

Himmeldirk, K. K., Sayer, B. G. & Spenser, I. D. (1998). *J. Am. Chem. Soc.* **120**, 3581–3589.

Hohmann, S. & Meacock, P. A. (1998). *Biochim. Biophys. Acta*, **1385**, 201–219.

Itou, H., Yao, M., Watanabe, N. & Tanaka, I. (2004). *Acta Cryst.* D**60**, 1094–1100.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). *Nature (London)*, **365**, 512–520.

Logan, D. T., Su, X. D., Aberg, A., Regnstrom, K., Hajdu, J., Eklund, H. & Nordlund, P. (1996). *Structure*, **4**, 1053–1064.

Morett, E., Korbel, J. O., Rajan, E., Saab-Rincon, G., Olvera, L., Olvera, M., Schmidt, S., Snel, B. & Bork, P. (2003). *Nature Biotechnol.* **21**, 790–795.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins*, **11**, 281–296.

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

Pang, A. S., Nathoo, S. & Wong, S. L. (1991). *J. Bacteriol.* **173**, 46–54.

Priest, F. G. (1977). *Bacteriol. Rev.* **41**, 711–753.

Risler, J. L., Delorme, M. O., Delacroix, H. & Henaut, A. (1988). *J. Mol. Biol.* **204**, 1019–1029.

Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. (2002). *J. Biol. Chem.* **277**, 48949–48959.

Schowen, R. L. (1998). *Comprehensive Biological Catalysis*, Vol 2, edited by M. Sinnott. New York: Academic Press.

Schuller, D. J., Wilks, A., Ortiz de Montellano, P. R. & Poulos, T. L. (1999). *Nature Struct. Biol.* **6**, 860–867.

Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). *Science*, **253**, 1001–1007.

Settembre, E., Begley, T. P. & Ealick, S. E. (2003). *Curr. Opin. Struct. Biol.* **13**, 739–747.

Stammers, D. K., Ren, J., Leslie, K., Nichols, C. E., Lamb, H. K., Cocklin, S., Dodds, A. & Hawkins, A. R. (2001). *EMBO J.* **20**, 6619–6626.

Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997). *Science*, **278**, 631–637.

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

Toms, A. V., Haas, A. L., Park, J.-H., Begley, T. P. & Ealick, S. E. (2005). *Biochemistry*, **44**, 2319–2329