

Gye Won Han,<sup>a,b</sup> Xiang-Lei Yang,<sup>b</sup>  
Daniel McMullan,<sup>a,c</sup> Yeeting E.  
Chong,<sup>b</sup> S. Sri Krishna,<sup>a,d,e</sup>  
Christopher L. Rife,<sup>a,f</sup> Dana Weekes,<sup>a,e</sup>  
Scott M. Brittain,<sup>c</sup> Polat Abdubek,<sup>a,c</sup>  
Eileen Ambing,<sup>a,c</sup> Tamara  
Astakhova,<sup>a,d</sup> Herbert L. Axelrod,<sup>a,f</sup>  
Dennis Carlton,<sup>a,b</sup> Jonathan  
Caruthers,<sup>a,f</sup> Hsiu-Ju Chiu,<sup>a,f</sup> Thomas  
Clayton,<sup>a,b</sup> Lian Duan,<sup>a,d</sup> Julie  
Feuerhelm,<sup>a,c</sup> Joanna C. Grant,<sup>a,c</sup>  
Slawomir K. Grzechnik,<sup>a,d</sup> Lukasz  
Jaroszewski,<sup>a,d,e</sup> Kevin K. Jin,<sup>a,f</sup>  
Heath E. Klock,<sup>a,c</sup> Mark W. Knuth,<sup>a,c</sup>  
Abhinav Kumar,<sup>a,f</sup> David Marciano,<sup>a,b</sup>  
Mitchell D. Miller,<sup>a,f</sup> Andrew T.  
Morse,<sup>a,d</sup> Edward Nigoghossian,<sup>a,c</sup>  
Linda Okach,<sup>a,c</sup> Jessica Paulsen,<sup>a,c</sup> Ron  
Reyes,<sup>a,f</sup> Henry van den Bedem,<sup>a,f</sup>  
Aprilfawn White,<sup>a,c</sup> Guenter Wolf,<sup>a,f</sup>  
Qingping Xu,<sup>a,f</sup> Keith O. Hodgson,<sup>a,g</sup>  
John Wooley,<sup>a,d</sup> Ashley M. Deacon,<sup>a,f</sup>  
Adam Godzik,<sup>a,d,e</sup> Scott A. Lesley,<sup>a,b,c</sup>  
Marc-André Elslinger,<sup>a,b</sup> Paul Schimmel<sup>b</sup>  
and Ian A. Wilson<sup>a,b\*</sup>

<sup>a</sup>Joint Center for Structural Genomics,  
http://www.jcsg.org, USA, <sup>b</sup>Department of  
Molecular Biology, The Scripps Research  
Institute, La Jolla, CA, USA, <sup>c</sup>Protein Sciences  
Department, Genomics Institute of the Novartis  
Research Foundation, San Diego, CA, USA,  
<sup>d</sup>Center for Research in Biological Systems,  
University of California, San Diego, La Jolla, CA,  
USA, <sup>e</sup>Program on Bioinformatics and Systems  
Biology, Sanford–Burnham Medical Research  
Institute, La Jolla, CA, USA, <sup>f</sup>Stanford  
Synchrotron Radiation Lightsource, SLAC  
National Accelerator Laboratory, Menlo Park,  
CA, USA, and <sup>g</sup>Photon Science, SLAC National  
Accelerator Laboratory, Menlo Park, CA, USA

Correspondence e-mail: wilson@scripps.edu

Received 27 August 2010

Accepted 20 September 2010

**PDB Reference:** tryptophanyl-tRNA synthetase  
from *T. maritima*, 2g36.

## Structure of a tryptophanyl-tRNA synthetase containing an iron–sulfur cluster

A novel aminoacyl-tRNA synthetase that contains an iron–sulfur cluster in the tRNA anticodon-binding region and efficiently charges tRNA with tryptophan has been found in *Thermotoga maritima*. The crystal structure of *TmTrpRS* (tryptophanyl-tRNA synthetase; TrpRS; EC 6.1.1.2) reveals an iron–sulfur [4Fe–4S] cluster bound to the tRNA anticodon-binding (TAB) domain and an L-tryptophan ligand in the active site. None of the other *T. maritima* aminoacyl-tRNA synthetases (AARSs) contain this [4Fe–4S] cluster-binding motif (C-x<sub>22</sub>-C-x<sub>6</sub>-C-x<sub>2</sub>-C). It is speculated that the iron–sulfur cluster contributes to the stability of *TmTrpRS* and could play a role in the recognition of the anticodon.

### 1. Introduction

Aminoacyl-tRNA synthetases (AARSs) covalently append amino acids to their cognate tRNAs. This reaction proceeds in two steps. The first involves the activation of the amino acid by ATP to form aminoacyl-adenylate, which then reacts with its corresponding tRNA to form aminoacyl-tRNA. All organisms possess separate tRNA synthetases for each of the 20 standard amino acids. AARSs are grouped into two classes (classes I and II; Eriani *et al.*, 1990) based on similarities in their sequences and structures and each group contains ten members. Class I AARSs are mostly monomeric and contain a classic Rossmann nucleotide-fold catalytic domain and two highly conserved sequence motifs ‘HIGH’ and ‘KMSKS’ that are critical for their function. Class II AARSs are structurally distinct from their class I counterparts; instead of the Rossmann fold they contain a central antiparallel  $\beta$ -sheet flanked by  $\alpha$ -helices and are mostly dimeric or multimeric. The reactions catalysed by the two classes differ: in class I AARSs the aminoacyl group is coupled to the 2'-hydroxyl of the tRNA, while in class II AARSs the 3'-hydroxyl is preferred. Tryptophanyl-tRNA synthetase (TrpRS; EC 6.1.1.2) belongs to the class I AARSs and, as its name implies, catalyzes the activation of tryptophan by ATP and the subsequent transfer of the tryptophanyl moiety onto the cognate tRNA.

Here, we report a novel TrpRS from *Thermotoga maritima* that contains a [4Fe–4S] cluster bound to the tRNA anticodon-binding (TAB) domain and an L-tryptophan located in the active site. The *TmTrpRS* structure was determined using the semi-automated high-throughput pipeline of the Joint Center for Structural Genomics (JCSG; Lesley *et al.*, 2002) as part of the National Institute of General Medical Sciences' Protein Structure Initiative (PSI).

### 2. Materials and methods

#### 2.1. Protein production and crystallization

*TM0492* (GenBank AAD35577.1; gi:4981003; Swiss-Prot Q9WYW2) was amplified by polymerase chain reaction (PCR) from *T. maritima* MSB8 genomic DNA using *PfuTurbo* (Stratagene) and primers (forward primer, 5'-TTGAGAATACTGAGCGGCATGAGACC; reverse primer, 5'-gagtttaattaattaGAACATCAGGTTTCATGGCCCTTCTCAC; target sequence in upper case) corresponding to the predicted 5' and 3' ends. The PCR product was cloned into plasmid pMH2T7, which encodes a noncleavable expression and purification tag (MGSDKIHSHHHH) at the amino-terminus of the full-length protein. The cloning junctions were confirmed by DNA

sequencing. Protein expression was performed in modified Terrific Broth using the *Escherichia coli* methionine-auxotrophic strain DL41. At the end of fermentation, lysozyme was added to the culture to a final concentration of 250  $\mu\text{g ml}^{-1}$  and the cells were harvested. After one freeze–thaw cycle, the cells were sonicated in lysis buffer [50 mM Tris pH 7.9, 50 mM NaCl, 10 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP)] and the lysate was clarified by centrifugation at 32 500g for 30 min. The soluble fraction was applied onto nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer, the resin was washed with wash buffer [50 mM Tris pH 7.9, 300 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP] and the protein was eluted with elution buffer [20 mM Tris pH 7.9, 300 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP]. The eluate was diluted tenfold with buffer Q [20 mM Tris pH 7.9, 5% (v/v) glycerol, 0.5 mM TCEP] containing 50 mM NaCl and loaded onto a RESOURCE Q column (GE Healthcare) pre-equilibrated with the same buffer. The protein was eluted with a linear gradient of 50–500 mM NaCl in buffer Q, buffer-exchanged with crystallization buffer [20 mM Tris pH 7.9, 150 mM NaCl, 0.5 mM TCEP] and concentrated to 18 mg ml<sup>-1</sup> by centrifugal ultrafiltration (Millipore) for crystallization assays. *TmTrpRS* was crystallized by mixing 200 nl protein solution with 200 nl crystallization solution and equilibrating against 50  $\mu\text{l}$  reservoir solution in the crystallization plate (Greiner Crystal Quick 96LP) using the nanodroplet vapor-diffusion method (Santarsiero *et al.*, 2002) with standard JCSG crystallization protocols (Lesley *et al.*, 2002). The expression and purification tag was not removed from the protein prior to crystallization. The crystallization reagent consisted of 12.5% (w/v) polyethylene glycol 3000, 0.25 M MgCl<sub>2</sub> and 0.1 M cacodylate pH 6.5. A plate-shaped crystal of approximate dimensions 0.15 × 0.10 × 0.02 mm was harvested after 18 d at 277 K for data collection. Glycerol was diluted to 20% (v/v) using the reservoir solution and then added in a 1:1 ratio to the drop as a cryoprotectant prior to mounting. Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM; Cohen *et al.*, 2002) at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, California, USA). The diffraction data were indexed in the orthorhombic space group *C*222<sub>1</sub> (Table 1). The molecular weight and oligomeric state of *TmTrpRS* in solution were determined using a 1 × 30 cm Superdex 200 column (GE Healthcare) coupled with miniDAWN static light-scattering (SEC/SLS) and Optilab differential refractive-index detectors (Wyatt Technology). The mobile phase consisted of 20 mM Tris pH 8.0, 150 mM NaCl and 0.02% (w/v) sodium azide.

## 2.2. Data collection, structure solution and refinement

Native diffraction data were collected on beamline 8.2.2 at the Advanced Light Source (ALS, Berkeley, USA). The data sets were collected at 100 K using an ADSC Quantum 315 CCD detector. Data were integrated and reduced using *XDS* and scaled with the program *XSCALE* (Kabsch, 1993, 2010a,b). The structure was determined using the JCSG molecular-replacement pipeline (Schwarzenbacher *et al.*, 2008) with TrpRS\_II from *Deinococcus radiodurans* (*DrTrpRS\_II*; Buddha & Crane, 2005b; PDB code 1yi8, chain B, sequence identity of 43%) as a search model. The initial solution was found using *MOLREP* (Vagin & Teplyakov, 1997). Model completion was performed with *Coot* (Emsley & Cowtan, 2004). TLS refinement was performed with *REFMAC5* (Winn *et al.*, 2003; three TLS groups; group 1, residues 0–112; group 2, residues 120–196; group 3, residues 197–328 and the [4Fe–4S] cluster) using a maximum-likelihood target function and individual *B*-factor refinement with appropriate

**Table 1**

Summary of crystal parameters, data-collection and refinement statistics for *TmTrpRS* (PDB code 2g36).

Values in parentheses are for the highest resolution shell.

Space group	<i>C</i> 222 <sub>1</sub>
Unit-cell parameters (Å)	<i>a</i> = 122.89, <i>b</i> = 152.73, <i>c</i> = 53.07
Data collection	$\lambda_1$
Wavelength (Å)	1.0000
Resolution range (Å)	29.64–2.50 (2.59–2.50)
No. of observations	63816
No. of reflections	17636
Completeness (%)	99.3 (97.9)
Mean <i>I</i> / $\sigma$ ( <i>I</i> )	10.9 (2.3)
<i>R</i> <sub>merge</sub> on <i>I</i> <sup>†</sup>	0.10 (0.64)
Model and refinement statistics	
Resolution range (Å)	29.64–2.50
No. of reflections (total)	17623 <sup>‡</sup>
No. of reflections (test)	894
Completeness (%)	99.3
Data set used in refinement	$\lambda_1$
Cutoff criterion	<i>F</i> <sub>1</sub>   > 0
<i>R</i> <sub>cryst</sub> <sup>§</sup>	0.19
<i>R</i> <sub>free</sub> <sup>¶</sup>	0.26
Stereochemical parameters	
Restraints (r.m.s.d., observed)	
Bond lengths (Å)	0.013
Bond angles (°)	1.60
Average isotropic <i>B</i> value, protein (Å <sup>2</sup> )	45.2
Average isotropic <i>B</i> value, ligands <sup>††</sup> (Å <sup>2</sup> )	58.7
Average isotropic <i>B</i> value, water (Å <sup>2</sup> )	41.3
ESU <sup>‡‡</sup> based on <i>R</i> <sub>free</sub> value (Å)	0.26
Protein residues/atoms	322/2563
Ligand/atoms	2/23
Solvent molecules	56

<sup>†</sup>  $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ . <sup>‡</sup> Typically, the number of unique reflections used in refinement is slightly less than the total number that were integrated and scaled. Reflections are excluded owing to systematic absences, negative intensities and rounding errors in the resolution limits and unit-cell parameters. <sup>§</sup>  $R_{\text{cryst}} = \frac{\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_{hkl} |F_{\text{obs}}|}$ , where  $F_{\text{calc}}$  and  $F_{\text{obs}}$  are the calculated and observed structure-factor amplitudes, respectively. <sup>¶</sup>  $R_{\text{free}}$  is the same as  $R_{\text{cryst}}$  but for 5.1% of the total reflections chosen at random and omitted from refinement. <sup>††</sup> L-Trp and iron-sulfur cluster. <sup>‡‡</sup> Estimated overall coordinate error (Collaborative Computational Project, Number 4, 1994; Cruickshank, 1999).

restraints. Residues 113–119 were disordered and were not refined. *CCP4* programs were used for data conversion and other calculations (Collaborative Computational Project, Number 4, 1994). Data-processing and refinement statistics are summarized in Table 1.

## 2.3. Validation and deposition

The quality of the crystal structure was analyzed using the *JCSG Quality Control* server (<http://smb.slac.stanford.edu/jcsg/QC>). This server processes the coordinates and data through a variety of validation tools including *AutoDepInputTool* (Yang *et al.*, 2004), *MolProbity* (Chen *et al.*, 2010), *WHAT IF* 5.0 (Vriend, 1990), *RESOLVE* (Terwilliger, 2003) and *MOLEMAN2* (Kleywegt, 2000), as well as several in-house scripts, and summarizes the results. Protein quaternary-structure analysis was performed with the *PQS* server (Henrick & Thornton, 1998). The sequence alignment was adapted from an analysis using *ClustalW* (Larkin *et al.*, 2007) and all other figures were prepared with *PyMOL* (DeLano Scientific). Atomic coordinates and experimental structure factors for *TmTrpRS* at 2.50 Å resolution have been deposited in the PDB (<http://www.pdb.org>) and are accessible under code 2g36.

## 2.4. Pyrophosphate-release assays

**2.4.1. Aminoacylation assay.** Aminoacylation assays were performed with 50 mM HEPES pH 7.5, 20 mM KCl, 10 mM MgCl<sub>2</sub>, 4 mM ATP, 1  $\mu\text{M}$  [<sup>3</sup>H]-L-Trp, 19  $\mu\text{M}$  L-Trp, 2 mM DTT and 160  $\mu\text{M}$  bulk *E. coli* tRNA. The aminoacylation reactions were initiated by

the addition of 100 nM enzyme pre-incubated at either 310 or 333 K. Samples were collected at various time points and quenched into a PVDF Multiscreen filter plate containing 100 mM EDTA, 300 mM sodium acetate pH 3.0 and 0.5 mg ml<sup>-1</sup> DNA as a carrier. Trichloroacetic acid was then added to each well at a 10% final concentration to precipitate the tRNA. The plate was then vacuum-dried and washed four times with cold wash solution (5% trichloroacetic acid and 100 mM cold L-Trp) to reduce the background radioactivity from free [<sup>3</sup>H]-L-Trp and once with 95% ethanol before scintillation counting.

**2.4.2. ATP-PP<sub>i</sub> exchange assay.** PP<sub>i</sub>-exchange reactions were performed in 100 mM HEPES pH 7.5, 20 mM KCl, 10 mM MgCl<sub>2</sub>, 2 mM ATP, 2 mM sodium PP<sub>i</sub>, [<sup>32</sup>P]-sodium PP<sub>i</sub>, 2 mM L-Trp and 5 mM β-mercaptoethanol. Reactions were initiated by the addition

of 1 μM enzyme pre-incubated at either 310 or 333 K. At each time point, samples were quenched into a PVDF Multiscreen filter plate containing 4% charcoal, 1 M HCl, 200 mM sodium PP<sub>i</sub>. The charcoal was collected and washed four times with 1 M HCl and 200 mM sodium PP<sub>i</sub> prior to scintillation counting.

2.5. Comparison with other tRNA synthetases

An iterative PSI-BLAST (Altschul *et al.*, 1997) search was performed for 20 rounds of three iterations each against the NCBI nonredundant (nr) protein-sequence database, using the tryptophanyl-tRNA synthetase sequence from *T. maritima* (gi:4981003) as the initial query. The resulting list of homolog sequences was then queried for the presence of the C-x<sub>6</sub>-C-x<sub>2</sub>-C pattern from the [4Fe-4S] cluster-binding motif. False-positive hits that contained the Cys pattern at a location other than the TAB region were discarded, resulting in 85 sequences that contained the cluster-binding motif. In addition, a PSI-BLAST filtered search using only the C-x<sub>6</sub>-C-x<sub>2</sub>-C pattern was performed using the sequence of *Tm*TrpRS as a query. An additional 22 unique sequences that were not identified by the previous method were found. Moreover, to ensure that we had exhaustively queried all proteins annotated as tryptophanyl-tRNA synthetases that contained the [4F-4S] cluster-binding motif, a text search was performed in which the nr database was mined for all annotations containing 'Tryptophan-tRNA synthetase' and variants thereof. The resulting sequences were then searched for the aforementioned motif of interest. An additional 104 sequences were found using this last approach that were not identified using the previous methods. Although most organisms possess only a single copy of each tRNA synthetase gene, two copies of TrpRS were detected in some bacterial species. The sequences were further analyzed to determine the distribution of the [4Fe-4S] cluster-binding motif. Alignment of the resulting sequences is shown in Supplementary Fig. S1<sup>1</sup>, where only one representative sequence from a clustering at 50% (sequences that are ≥50% identical are clustered as single representative) is shown.

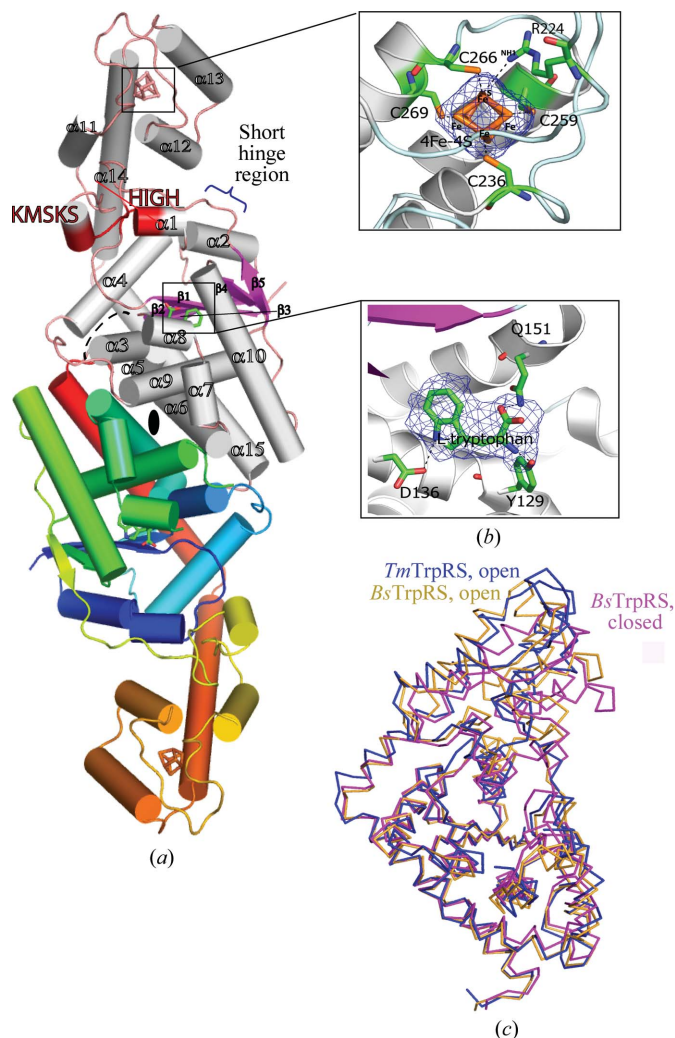
3. Results and discussion

3.1. Overall structure description

The 2.5 Å resolution crystal structure of *Tm*TrpRS was determined by molecular replacement (MR) using *Dr*TrpRS\_II (PDB code 1yi8; Buddha & Crane, 2005b) as a search model. The final model contains one *Tm*TrpRS molecule (residues 0–328), one iron-sulfur cluster [4Fe-4S], one tryptophan and 56 water molecules in the asymmetric unit (Fig. 1). No interpretable electron density was observed for residues -11 through -1 (from the expression and purification tag) and residues 113–119. The side chains of Lys12, Lys27, Glu74, Thr108, Lys110, Glu111, Glu149, Asp150, Arg182, Lys185, Lys193, Lys209, Glu210, Gu212, Gln213, Lys253, Glu257, Lys271, Lys307 and Lys315 were omitted owing to poor electron density. The Matthews coefficient ( $V_M$ ; Matthews, 1968) is 3.49 Å<sup>3</sup> Da<sup>-1</sup> and the estimated solvent content is 64.8%. The Ramachandran plot produced by *MolProbity* shows that 97% of the residues are in favored regions, with no outliers.

The *Tm*TrpRS crystal structure, like other bacterial TrpRS structures, contains two domains. The N-terminal catalytic domain adopts a canonical Rossmann nucleotide-binding fold (residues 1–181 and 296–328) with a central, five-stranded, parallel β-sheet (β1–β5; the

<sup>1</sup> Supplementary material has been deposited in the IUCr electronic archive (Reference: WD5146).



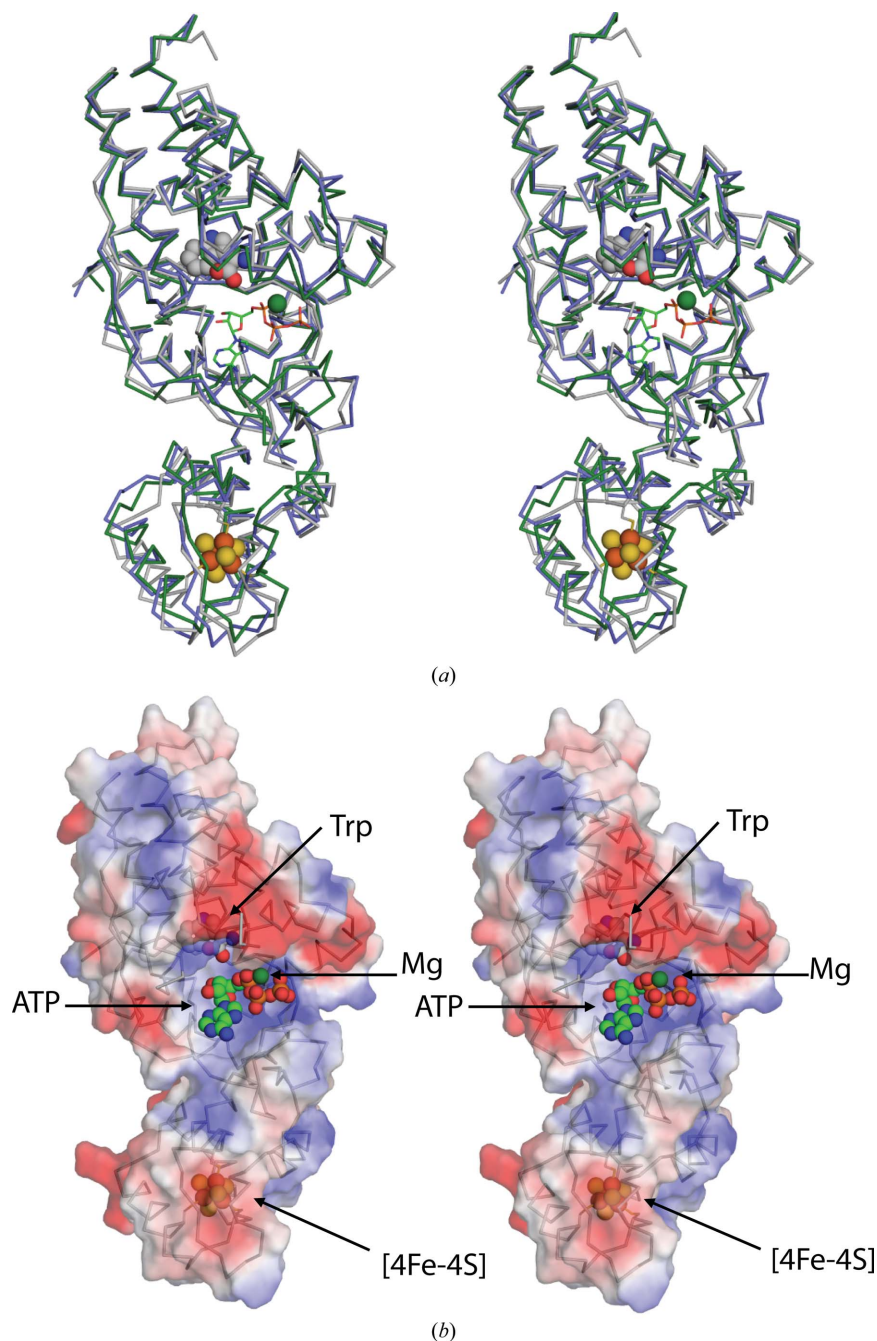
**Figure 1** Crystal structure of *Tm*TrpRS. (a) Schematic representation of the *Tm*TrpRS crystallographic dimer. The α-helices are represented as cylinders in gray, β-strands as arrows in purple and the loops in pink in one monomer (top) of the dimer. The crystallographically related molecule below is color-coded from blue at the N-terminus to red at the C-terminus. The bound 4Fe-4S cluster and L-tryptophan are represented as sticks. The β-strands (β1–β5) and α-helices (α1–α15) are labeled (β<sub>10</sub>-helices are not labeled). The ATP-binding motifs (HIGH and KMSKS) are highlighted in red. The short hinge region (residues 182–186) connecting the Rossmann-fold and the TAB domains is indicated. (b) Close-up view of a 2F<sub>o</sub> – F<sub>c</sub> OMIT map contoured at 1σ showing the iron-sulfur cluster bound to the cysteine motif [Cys236-x<sub>22</sub>-Cys259-x<sub>6</sub>-Cys266-x<sub>2</sub>-Cys269] in the anticodon-binding region (top box) and the L-tryptophan found in the active site (bottom box). (c) Comparison of *Tm*TrpRS to the 'open' and 'closed' conformations of *Bs*TrpRS.

strand order is 32145) and a tRNA anticodon-binding (TAB) domain that adopts an all-helical fold (residues 187–294) (Fig. 1). The TAB domain is composed of four  $\alpha$ -helices ( $\alpha$ 11– $\alpha$ 14) and two  $3_{10}$  helices ( $\eta$ 3 and  $\eta$ 4) that are packed as a bundle. A short hinge region (residues 182–186; Fig. 1a) connects the Rossmann-fold and the TAB domains. The asymmetric unit contains one monomer of *Tm*TrpRS, which forms a crystallographic dimer across the twofold.

As per the SCOP classification (Murzin *et al.*, 1995), the catalytic domain of TrpRSs belongs to the nucleotidyl transferase superfamily. While all members of this superfamily retain the core

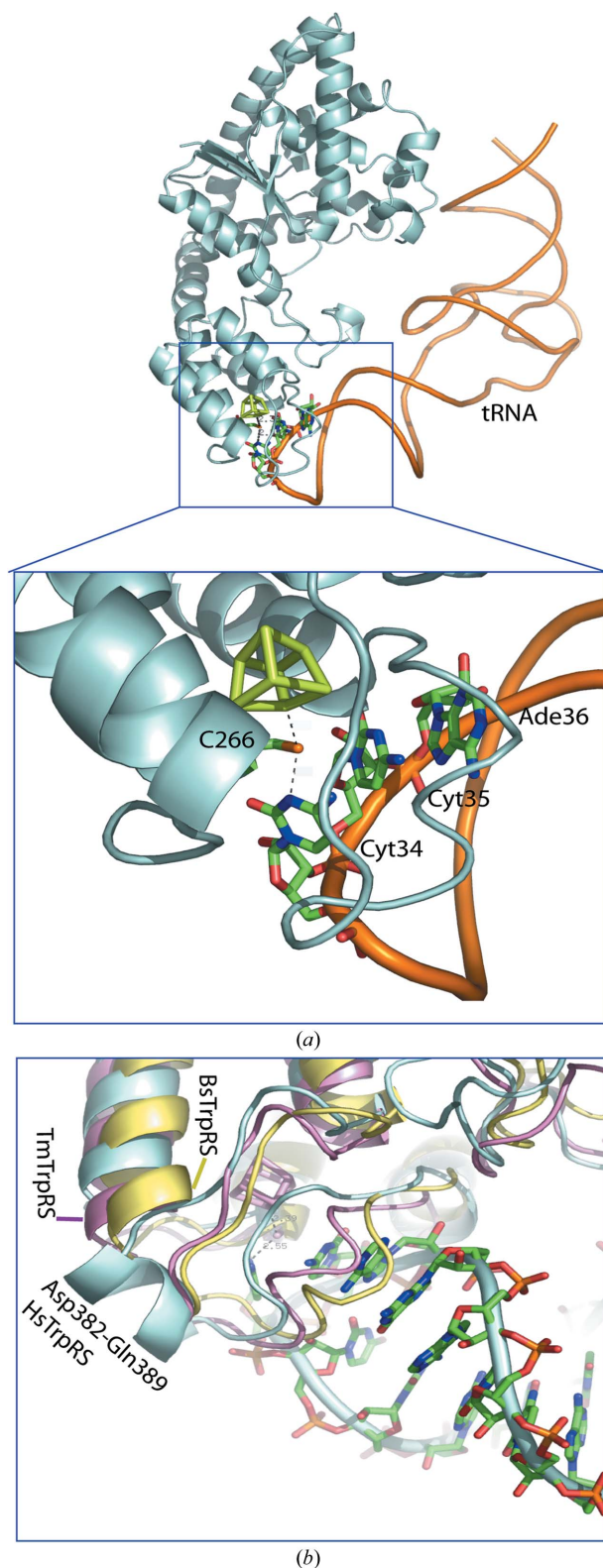
elements of the Rossmann fold, substantial insertions to the catalytic domain, which confer novel functions, have been observed.

As expected, a structure-similarity search using *DALI* (Holm & Sander, 1995) revealed extensive similarities to several class I AARSs. Top *DALI* hits include TrpRS\_II from *D. radiodurans* (*Dr*TrpRS\_II; Buddha & Crane, 2005a; PDB code 1yid;  $Z = 39.7$ ), TrpRS from *Bacillus stearothermophilus* (*Bs*TrpRS; Ilyin *et al.*, 2000; PDB code 1d2r;  $Z = 22.5$ ) and human tyrosyl-tRNA synthetase (*Hs*TyrRS; Yang *et al.*, 2003; PDB code 1q11;  $Z = 20.3$ ), among many other class I AARSs. The structural similarity between *Dr*TrpRS-II



**Figure 2**

Comparison of *Tm*TrpRS and *Dr*TrpRS\_II. (a) Structural superposition of *Tm*TrpRS (gray), *Dr*TrpRS\_II + ATP and Mg (PDB code 1yid; green) and *Dr*TrpRS\_II + Trp (PDB code 1yi8; blue). The bound ATP is shown as a stick model. The [4Fe-4S] cluster (orange and yellow), Mg ion (green sphere) and Trp (gray, red and blue) are shown as spheres. The *Dr*TrpRS\_II structures superimpose with *Tm*TrpRS over 319 C $\alpha$  atoms with an r.m.s.d. of  $\sim 1.7$  Å (45% sequence identity). (b) Electrostatic surface potential of *Tm*TrpRS calculated with the program *APBS* (Baker *et al.*, 2001): positive potential is shown in blue ( $+3kTe^{-1}$ ) and negative in red ( $-3kTe^{-1}$ ). L-Tryptophan and the iron-sulfur cluster are shown as spheres. Note: the ATP-binding site is solvent-exposed, but the tryptophan and the iron-sulfur cluster are partially buried.



**Figure 3**  
 Model of the *Tm*TrpRS-tRNA complex. (a) Overall view of the complex. *Tm*TrpRS is shown as a gray-blue ribbon and tRNA as an orange trace. The anticodon bases, Cys266 and the [4Fe-4S] cluster are shown as stick models. Close-up view of the interaction of TrpRS-tRNA with the anticodon (CCA) region. Cyt34 of the tRNA molecule interacts with the iron of the cluster via Cys266. The model was based on the structure of the human TrpRS-tRNA complex (PDB code 1r6t). (b) Close-up view of the loop conformation of *Tm*TrpRS (pink) near the [4Fe-4S] cluster (pink sticks) compared with the helix (Asp382-Asn389) of the human TrpRS-tRNA complex (gray-blue) and *Bs*TrpRS (yellow).

and *Tm*TrpRS is particularly high and a superimposition of the structures based on secondary-structural elements from both the catalytic Rossmann-fold domain and the C-terminal TAB domain gives an r.m.s.d. of 1.7 Å (Fig. 2).

Remarkably, the structure of *Tm*TrpRS differs from other TrpRS by the presence of an iron-sulfur cluster [4Fe-4S] in the C-terminal TAB domain. In addition, an L-tryptophan molecule is bound in the active site, which was not expected since tryptophan was not added to any of the reagents used in crystallization or purification.

### 3.2. [4Fe-4S] cluster-binding site

The [4Fe-4S] cluster is chelated by the side chains of Cys236, Cys259, Cys266 and Cys269 from the TAB domain arranged in a C-x<sub>22</sub>-C-x<sub>6</sub>-C-x<sub>2</sub>-C motif in which four irons are bound to the S atoms of cysteines with distances of 2.28–2.34 Å. The other significant interaction involves the NH1 atom of Arg224 and one of the S atoms (S2) from the cluster, with a distance of 3.31 Å (Fig. 1b). The presence of the [4Fe-4S] cluster was first identified based on electron density and geometry. The presence of iron in the structure was then confirmed by X-ray fluorescence scans (Supplementary Fig. S2; for details of how ligands are identified at the JCSG in the course of structure determination, see Kumar *et al.*, 2010). Mass spectrometry also corroborated the presence of the iron-sulfur cluster. Although *Tm*TrpRS shares extensive sequence similarity in the TAB region with other TrpRSs, the C-x<sub>22</sub>-C-x<sub>6</sub>-C-x<sub>2</sub>-C motif is not found in any other TrpRS present in the PDB. Sequences of *Tm*TrpRS homologs that possess the [4Fe-4S] cluster-binding motif are found in anaerobic organisms from proteobacterial and archaeal groups, but no structures of any of these have yet been reported. *Tm*TrpRS is thus the first reported structure of a TrpRS that contains a [4Fe-4S] cluster.

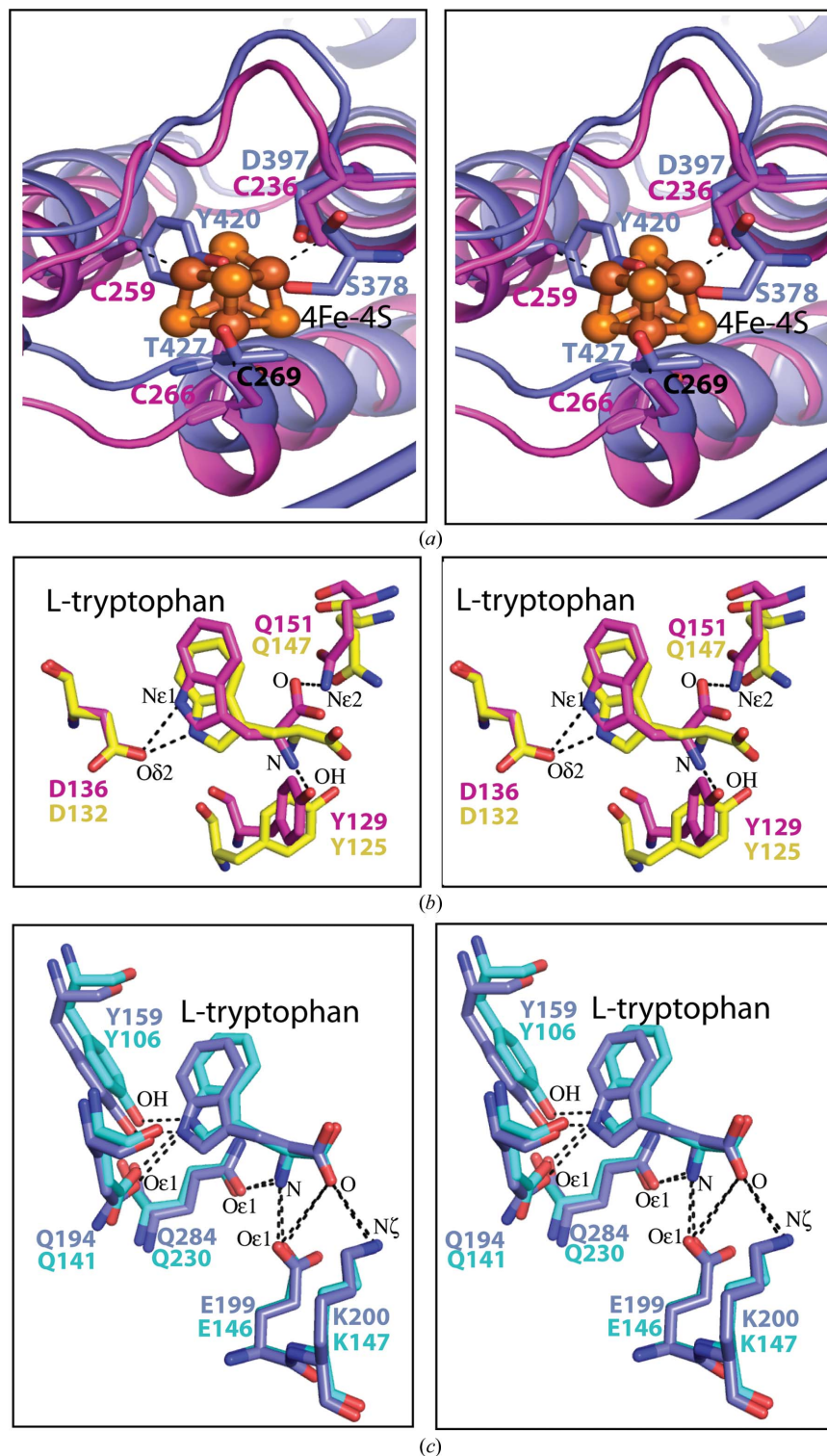
We extended our search to find other potential iron-sulfur clusters based on the presence of the cysteine-binding motif using *SPASM* (Madsen & Kleywegt, 2002). For this search, the coordinates of the four cysteines (Cys236, Cys259, Cys266 and Cys269) were used and no substitutions of amino acids were allowed. *SPASM* identified 71 hits representing 22 unique proteins within a 1.5 Å r.m.s.d. of the target motif. The hits included the DNA-repair enzyme endonuclease III (PDB code 2abk; Thayer *et al.*, 1995), acetyl-CoA synthase (PDB code 1ru3; Svetlitchnyi *et al.*, 2004) and carbon monoxide dehydrogenase (PDB code 1jqk; Drennan *et al.*, 2001). The top hit containing the closest structural homolog was the *E. coli* DNA glycosylase MutY (Guan *et al.*, 1998; PDB code 1mun), which belongs to the DNA-repair enzyme superfamily and excises adenine from mispairs with 8-oxoguanine and guanine. Although the [4Fe-4S] cluster-binding motif (C-x<sub>6</sub>-C-x<sub>2</sub>-C-x<sub>5</sub>-C) of *E. coli* MutY has, in particular, a smaller sequence gap between the first two Cys residues than *Tm*TrpRS (C-x<sub>22</sub>-C-x<sub>6</sub>-C-x<sub>2</sub>-C), the r.m.s.d. between the two motifs was only 0.88 Å.

Comparison of the *Tm*TrpRS structure with that of the human TrpRS-tRNA complex (PDB code 1r6t; Yang *et al.*, 2006) reveals that, although the sequence identity is very low (19%), the structures are very similar, with an r.m.s.d. of 1.9 Å for 238 superimposed C<sup>α</sup> atoms. A model of a *Tm*TrpRS-tRNA complex based on the complex of human TrpRS with tRNA<sup>Trp</sup> indicates that Cyt34 of the CCA anticodon of tRNA<sup>Trp</sup> can interact with the iron of the cluster via Cys266 (Fig. 3a).

Interestingly, the *T. maritima* tRNA-modifying enzyme MiaB (*Tm*MiaB), which is involved in the post-transcriptional thiolation and methylation of tRNA, contains an iron-sulfur cluster with a (C-x<sub>3</sub>-C-x<sub>2</sub>-C) binding motif (Pierrel *et al.*, 2003). The iron-sulfur

cluster in this case is essential for the modification of the tRNA adenine 37, which helps to stabilize the tRNA anticodon loop. This reaction is catalyzed by the iron that is not coordinated to a cysteine, which is a general theme for iron–sulfur clusters involved in catalysis. In the case of the [4Fe–4S] cluster in *Tm*TrpRS, we believe that it plays a role in the recognition of specifically modified tRNA;

however, the biological implications of this are unknown at present. Modifications of the nucleotide in the wobble position 34 are common (Gustilo *et al.*, 2008); for example, *Saccharomyces cerevisiae* mitochondrial tRNA<sup>Leu</sup> and tRNA<sup>Trp</sup> contain a modified U at the wobble position 34 (Martin *et al.*, 1990). Moreover, 2-thiocytidine is often found in the anticodon loop and all tRNA<sup>Arg</sup> species from



**Figure 4**

Iron–sulfur cluster and L-tryptophan recognition of *Tm*TrpRS and comparison with other TrpRSs. (a) Comparison of the *Tm*TrpRS (magenta) and human TrpRS (blue) TAB domains. The *Tm*TrpRS iron–sulfur cluster is shown in orange ball-and-stick representation. (b) Comparison of the L-tryptophan recognition of *Tm*TrpRS (magenta) and *Bs*TrpRS (yellow). (c) Comparison of the L-tryptophan recognition of human TrpRS (blue) and *Sc*TrpRS (slate blue).

*E. coli* (Jäger *et al.*, 2004), for example, contain a 2-thiocytidine at position 32.

An interesting difference in the *Tm*TrpRS structure compared with the human TrpRS–tRNA complex is the substitution of helix  $\alpha 17$  (Asp382–Gln389) in the human enzyme with a loop in *Tm*TrpRS (Fig. 3*b*). As revealed in the crystal structure of the human TrpRS–tRNA complex,  $\alpha 17$  is involved in the recognition of the anticodon of the tRNA (Yang *et al.*, 2006). A similar conformational change has been observed between *Bs*TrpRS and the human enzyme (Fig. 3*b*). It was suggested that tRNA binding to *Bs*TrpRS may induce the human enzyme-like conformation in this region (Yang *et al.*, 2006). Interestingly, the iron cluster is located near this loop. A *PSI-BLAST*

search for homologs of *Tm*TrpRS in the NCBI nonredundant (nr) protein-sequence database shows that the C-x<sub>(21–24)</sub>-C-x<sub>6</sub>-C-x<sub>2</sub>-C motif is mostly found in thermophiles or other extremophiles (Supplementary Fig. 1). Interestingly, this feature is found in organisms that possess either a single TrpRS gene or multiple genes encoding TrpRS. In those organisms that contain multiple TrpRS genes, only one copy contains the [4Fe–4S] cluster-binding motif (Supplementary Fig. S1).

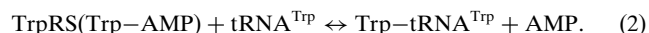
### 3.3. ATP- and Trp-binding sites

*Tm*TrpRS possesses ATP-binding and Trp-binding sites, which are located close to each other in the Rossmann-fold domain. Typically, this enzyme, which is an obligate dimer, binds ATP and L-tryptophan in one subunit, while the tRNA anticodon region is recognized by the TAB domain from the other subunit. Most class I AARSs are functional as monomers, with the exceptions of TrpRS and TyrRS, which are obligate homodimers. In *Tm*TrpRS, the two relevant subunits pack against each other, burying 2365 Å<sup>2</sup> of mainly hydrophobic surface. Although a single molecule is present in the asymmetric unit, crystal-packing analysis identified a crystallographic dimer that is likely to represent the biologically relevant dimer. Analytical size-exclusion chromatography in combination with static light scattering indicated that the major species in solution is a dimer.

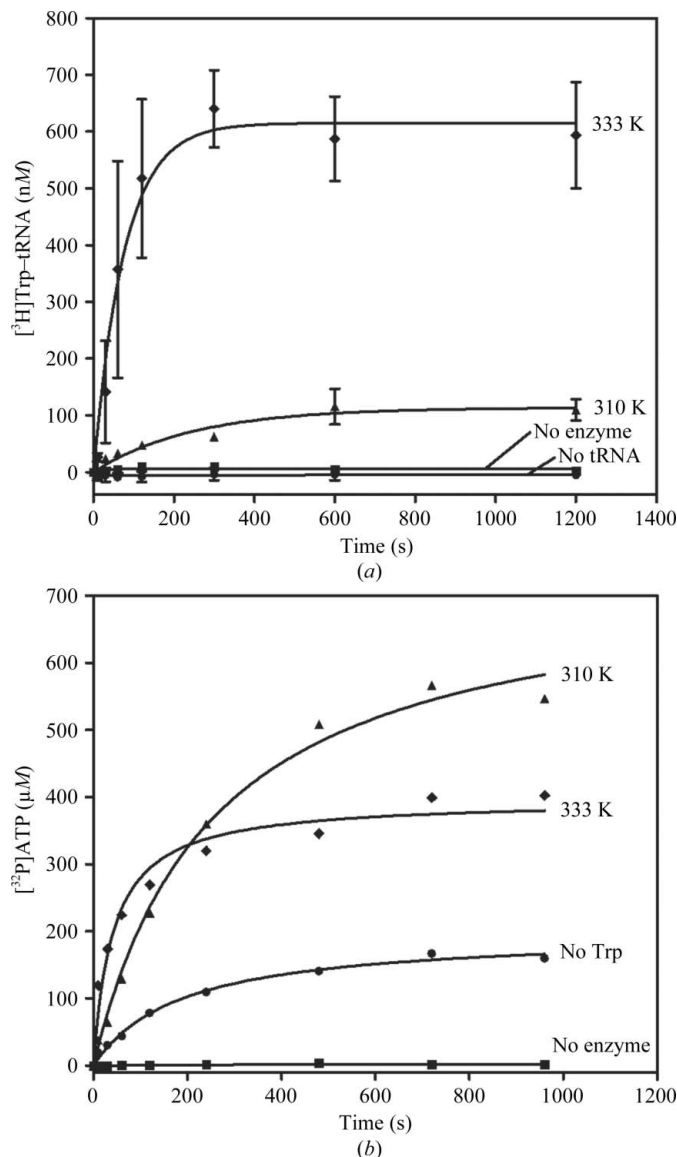
Although Trp was not present in any of the crystallization reagents, the structure revealed a bound L-tryptophan molecule in the Trp-binding pocket of the active site (Fig. 1*a*), suggestive of tight binding of *Tm*TrpRS towards the substrate. The L-tryptophan-binding site in the *Tm*TrpRS structure is similar to that seen in the *Bs*TrpRS and human TrpRS structures (Figs. 4*b* and 4*c*; see also §3.4), as are the relative orientations of the bound L-tryptophan. However, the L-tryptophan recognition in *Tm*TrpRS is more akin to that of *Bs*TrpRS than to that of human TrpRS.

The ATP-binding site is located in a positively charged, solvent-exposed cleft located at the junction of the two domains (Fig. 2). It contains the two signature sequences that are conserved across all members of the class I AARSs: the <sup>14</sup>HIGH<sup>17</sup> and <sup>193</sup>KMSKS<sup>197</sup> motifs responsible for binding to the adenosine moiety of ATP (Fig. 1*a*). The ATP-binding cleft opens and closes *via* a rotation about a hinge between the Rossmann-fold domain and the anticodon-recognition domain (Fig. 1*c*). In the *Tm*TrpRS structure, which is similar to the ‘open’ conformation of *Bs*TrpRS, the KMSKS motif is further from the ATP site and is not poised to bind ATP. Accordingly, no ATP molecule was observed in the crystal.

TrpRS activity was confirmed for TM0492. The activity was characterized in tryptophan-dependent ATP–PP<sub>i</sub> exchange (1) and aminoacylation assays (sum of equations 1 and 2),



The overall aminoacylation of tRNA<sup>Trp</sup> can be measured by the incorporation of L-tryptophan into the tRNA to form Trp–tRNA<sup>Trp</sup> in the presence of ATP. Consistent with its thermophilic nature, *Tm*TrpRS has a more robust tRNA-charging activity at 333 K compared with that at 310 K (Fig. 5*a*). The ATP–PP<sub>i</sub> exchange reaction assesses the reverse of amino-acid activation by measuring the incorporation of [<sup>32</sup>P]–PP<sub>i</sub> into ATP (1). TM0492 was also active in this assay at both 310 and 333 K (Fig. 5*b*). Although the amount of [<sup>32</sup>P]–ATP reached a lower plateau at 333 K than at 310 K, presumably owing to the increase in ATP hydrolysis at higher temperature, the initial PP<sub>i</sub>-exchange rate was higher at 333 K than at 310 K,



**Figure 5** Enzymatic activities of *Tm*TrpRS. (a) Aminoacylation activity assayed at 310 and 333 K. Consistent with its thermophilic nature, *Tm*TrpRS has a more robust tRNA-charging activity at 333 K compared with that at 310 K. Control reaction assays lacking enzyme or tRNA at 333 K are also shown. Points are the mean of two assays and error bars represent the standard error of the mean of measurements. (b) ATP–PP<sub>i</sub>-exchange activities assayed at 310 and 333 K. This experiment was only performed once. Plots were derived by fitting to an exponential rise to maximum function. Control reaction assays lacking enzyme or Trp at 333 K are also shown. Consistent with the observation of an endogenously bound Trp molecule in the active site of its crystal structure, *Tm*TrpRS had some PP<sub>i</sub>-exchange activity even when no Trp was added to the reaction.

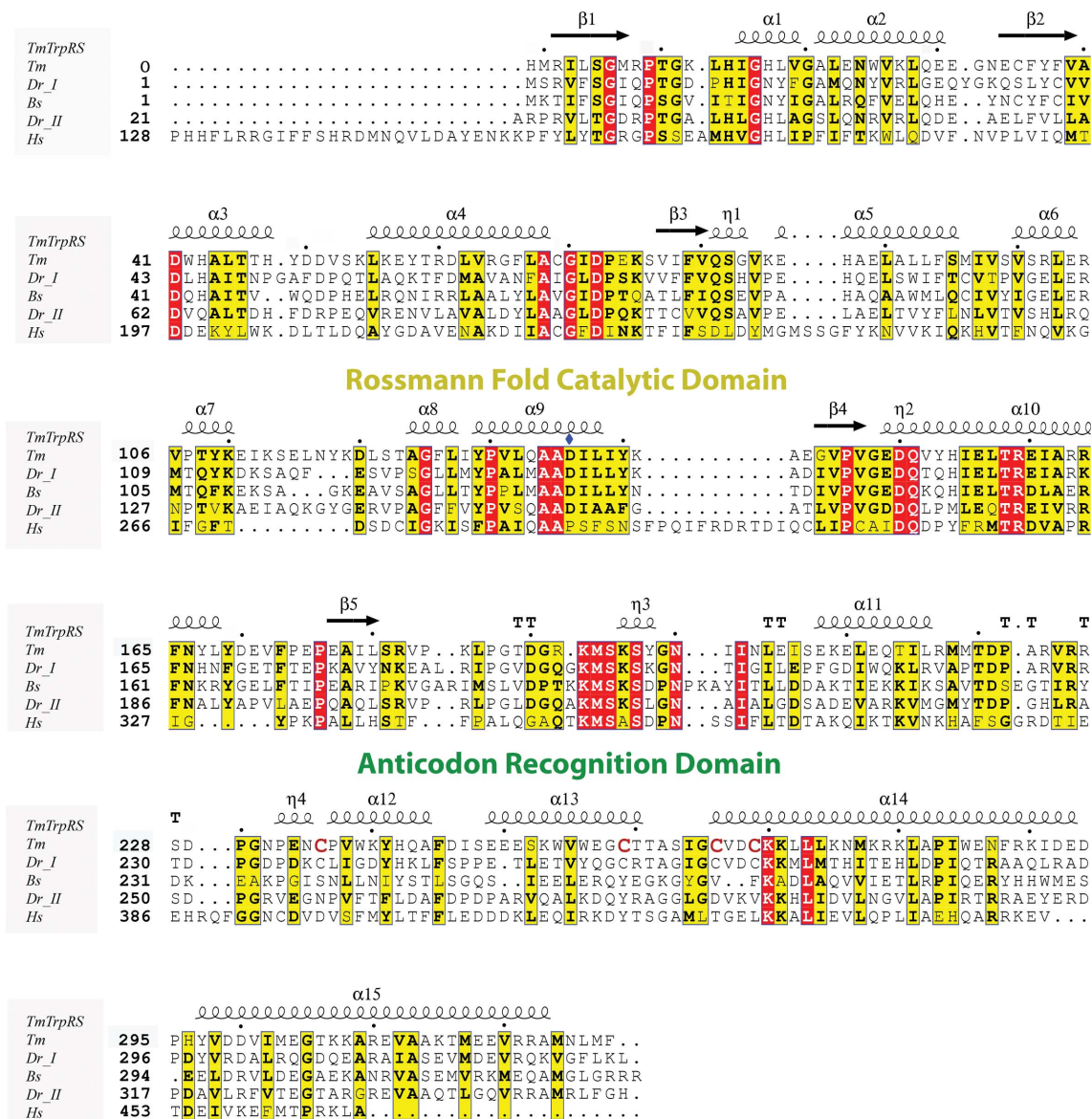
consistent with the thermophilic nature of TM0492. Furthermore, PP<sub>i</sub>-exchange activity was apparent even when no Trp was added to the reaction (Fig. 4b). This result is consistent with the observation of an endogenously bound tryptophan molecule in the active site of the crystal structure.

### 3.4. Ligand-binding modes in *Tm*TrpRS and human TrpRS

Structural comparison of the iron–sulfur cluster region of *Tm*TrpRS with that of human TrpRS shows that the chemical environment in this region is very different. There is nothing in the human enzyme that replaces the iron–sulfur cluster. The four-cysteine motif of *Tm*TrpRS, Cys236, Cys259, Cys266 and Cys269, is not conserved in the human enzyme (Fig. 4a). The corresponding residues in the human enzyme are Asp397, Tyr420, Thr427 and Leu430, respectively

(Fig. 6). In addition, the side chains of Tyr420 and Thr427 overlap with the iron–sulfur cluster of *Tm*TrpRS (Fig. 4a).

Structural comparison of the L-tryptophan-binding pockets of *Tm*TrpRS and human TrpRS revealed that the orientations and relative positions of the L-tryptophans are similar, but the interactions between Trp and the binding-pocket residues differ. N<sup>ε1</sup> of the tryptophan interacts with Asp136 in *Tm*TrpRS (Fig. 4b). This Asp is conserved among prokaryotes (Fig. 6). In human TrpRS, N<sup>ε1</sup> of the tryptophan interacts with Tyr159 and Gln194 (Fig. 4c). These Tyr and Gln residues are conserved in *S. cerevisiae* TrpRS as shown in the recent structure (PDB code 3kt0; Zhou *et al.*, 2010). L-Tryptophan recognition in TrpRS is highly conserved among eukaryotes (Zhou *et al.*, 2010) and the interactions observed in *Tm*TrpRS are conserved among prokaryotes (Figs. 4b and 4c).



**Figure 6** Structure-based sequence alignment of TrpRSs. The four Cys residues in the C-x<sub>22</sub>-C-x<sub>6</sub>-C-x<sub>2</sub>-C motif of *Tm*TrpRS that chelate the [4Fe–4S] cluster are colored by red letters. Strictly conserved residues are highlighted in red boxes and additional conserved residues are in yellow boxes. The secondary-structure elements of *Tm*TrpRS are shown at the top, where α-helices (α1–α15), β-strands (β1–β5) and 3<sub>10</sub>-helices (η1–η4) are sequentially labeled and β-turns and γ-turns are designated with the Greek letter tau (T). Asp136, which is conserved among prokaryotes, is indicated with a blue diamond. Abbreviations: *Thermotoga maritima* TrpRS, *Tm*TrpRS; *Deinococcus radiodurans* TrpRS\_I, *Dr\_I*; *Bacillus stearothermophilus* TrpRS, *Bs*; *D. radiodurans* TrpRS\_II, *Dr\_II*; *Homo sapiens* TrpRS, *Hs*. Note: the sequence of *Dr*TrpRS\_I is included in the alignment although there is no crystal structure available from the PDB.



## 4. Conclusions

We report the first structure of an iron–sulfur cluster-containing tRNA synthetase. Interestingly, this structure also revealed an L-tryptophan in the active site. The iron–sulfur cluster located in the anticodon-binding region is coordinated by a four-cysteine motif C-x<sub>22</sub>-C-x<sub>6</sub>-C-x<sub>2</sub>-C. The role of the iron–sulfur cluster is still not clear. The complexity and energetic cost of synthesizing a [4Fe–4S] cluster suggests that it may not be limited to performing a purely structural role.

In a model of the *Tm*TrpRS–tRNA complex based on the structure of the human complex, the [4Fe–4S] cluster is within contact distance of the tRNA anticodon (Fig. 3a). This implies that the iron–sulfur cluster could be crucial for anticodon recognition by *Tm*TrpRS. One hypothesis is that the [4Fe–4S] cluster could be involved in recognition of a modified tRNA anticodon. However, the modification state of the tRNA<sup>Trp</sup> anticodon of *Tm*TrpRS–tRNA is not known.

Availability of more sequences and structures of [4Fe–4S] cluster proteins might shed light on the evolutionary relationships of this enzyme. The information presented here, in combination with further biochemical and biophysical studies, should yield valuable insights into the functional role of the enzyme. Additional information about TM0492/*Tm*TrpRS described in this study is available from TOPSAN (Krishna *et al.*, 2010; Weekes *et al.*, 2010) at <http://www.topsan.org/explore?PDBid=2g36>.

This work was supported by the NIH, National Institute of General Medical Sciences, Protein Structure Initiative grant U54 GM074898. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL) and the Advanced Light Source (ALS). The SSRL is a national user facility operated by Stanford University on behalf of the US DOE, OBES. The SSRL Structural Molecular Biology Program is supported by the DOE, OBER and by NIH (NCRR, BTP and NIGMS). The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences of the US Department of Energy under Contract No. DE-AC02-05CH11231 at Lawrence Berkeley National Laboratory. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences. Genomic DNA from *T. maritima* MSB8 (DSM3109; ATCC #43589D-5) was obtained from the American Type Culture Collection (ATCC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

## References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.  
 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.  
 Buddha, M. R. & Crane, B. R. (2005a). *J. Biol. Chem.* **280**, 31965–31973.  
 Buddha, M. R. & Crane, B. R. (2005b). *Nature Struct. Mol. Biol.* **12**, 274–275.  
 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. D66*, 12–21.

Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). *J. Appl. Cryst.* **35**, 720–726.  
 Collaborative Computational Project, Number 4 (1994). *Acta Cryst. D50*, 760–763.  
 Cruickshank, D. W. J. (1999). *Acta Cryst. D55*, 583–601.  
 Drennan, C. L., Heo, J., Sintchak, M. D., Schreiter, E. & Ludden, P. W. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 11973–11978.  
 Emsley, P. & Cowtan, K. (2004). *Acta Cryst. D60*, 2126–2132.  
 Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990). *Nature (London)*, **347**, 203–206.  
 Guan, Y., Manuel, R. C., Arvai, A. S., Parikh, S. S., Mol, C. D., Miller, J. H., Lloyd, S. & Tainer, J. A. (1998). *Nature Struct. Biol.* **5**, 1058–1064.  
 Gustilo, E. M., Vendeix, F. A. & Agris, P. F. (2008). *Curr. Opin. Microbiol.* **11**, 134–140.  
 Henrick, K. & Thornton, J. M. (1998). *Trends Biochem. Sci.* **23**, 358–361.  
 Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.  
 Ilyin, V. A., Temple, B., Hu, M., Li, G., Yin, Y., Vachette, P. & Carter, C. W. Jr (2000). *Protein Sci.* **9**, 218–231.  
 Jäger, G., Leipuviene, R., Pollard, M. G., Qian, Q. & Bjork, G. R. (2004). *J. Bacteriol.* **186**, 750–757.  
 Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.  
 Kabsch, W. (2010a). *Acta Cryst. D66*, 125–132.  
 Kabsch, W. (2010b). *Acta Cryst. D66*, 133–144.  
 Kleywegt, G. J. (2000). *Acta Cryst. D56*, 249–265.  
 Krishna, S. S., Weekes, D., Bakolitsa, C., Elsliger, M.-A., Wilson, I. A., Godzik, A. & Wooley, J. (2010). *Acta Cryst. F66*, 1143–1147.  
 Kumar, A., Chiu, H.-J., Axelrod, H. L., Morse, A., Elsliger, M.-A., Wilson, I. A. & Deacon, A. (2010). *Acta Cryst. F66*, 1309–1316.  
 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.  
 Lesley, S. A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 11664–11669.  
 Madsen, D. & Kleywegt, G. J. (2002). *J. Appl. Cryst.* **35**, 137–139.  
 Martin, R. P., Sibley, A. P., Gehrke, C. W., Kuo, K., Edmonds, C. G., McCloskey, J. A. & Dirheimer, G. (1990). *Biochemistry*, **29**, 956–959.  
 Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.  
 Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.  
 Pierrel, F., Hernandez, H. L., Johnson, M. K., Fontecave, M. & Atta, M. (2003). *J. Biol. Chem.* **278**, 29515–29524.  
 Santarsiero, B. D., Yegian, D. T., Lee, C. C., Spraggon, G., Gu, J., Scheibe, D., Uber, D. C., Cornell, E. W., Nordmeyer, R. A., Kolbe, W. F., Jin, J., Jones, A. L., Jaklevic, J. M., Schultz, P. G. & Stevens, R. C. (2002). *J. Appl. Cryst.* **35**, 278–281.  
 Schwarzenbacher, R., Godzik, A. & Jaroszewski, L. (2008). *Acta Cryst. D64*, 133–140.  
 Svetlitchnyi, V., Dobbek, H., Meyer-Klaucke, W., Meins, T., Thiele, B., Romer, P., Huber, R. & Meyer, O. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 446–451.  
 Terwilliger, T. C. (2003). *Acta Cryst. D59*, 1174–1182.  
 Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P. & Tainer, J. A. (1995). *EMBO J.* **14**, 4108–4120.  
 Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.  
 Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.  
 Weekes, D., Krishna, S. S., Bakolitsa, C., Wilson, I. A., Godzik, A. & Wooley, J. (2010). *BMC Bioinformatics*, **11**, 426.  
 Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.  
 Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). *Acta Cryst. D60*, 1833–1839.  
 Yang, X.-L., Otero, F. J., Ewalt, K. L., Liu, J., Swairjo, M. A., Kohrer, C., RajBhandary, U. L., Skene, R. J., McRee, D. E. & Schimmel, P. (2006). *EMBO J.* **25**, 2919–2929.  
 Yang, X.-L., Otero, F. J., Skene, R. J., McRee, D. E., Schimmel, P. & Ribas de Pouplana, L. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 15376–15380.  
 Zhou, M., Dong, X., Shen, N., Zhong, C. & Ding, J. (2010). *Nucleic Acids Res.* **38**, 3399–3413.