

The folate-binding module of *Thermus thermophilus* cobalamin-dependent methionine synthase displays a distinct variation of the classical TIM barrel: a TIM barrel with a ‘twist’

Kazuhiro Yamada^{a,b} and Markos Koutmos^{a,b*}

Received 6 October 2017

Accepted 21 December 2017

^aDepartment of Chemistry, University of Michigan, Ann Arbor, Michigan, USA, and ^bProgram in Biophysics, University of Michigan, Ann Arbor, Michigan, USA. *Correspondence e-mail: mkoutmos@umich.edu

Edited by G. J. Kleywegt, EMBL-EBI, Hinxton, England

Keywords: methionine synthase; folate; pterin-binding proteins; TIM barrel; cobalamin.

PDB references: methionine synthase folate-binding domain, 5von; complex with methyltetrahydrofolate, 5voo; 5vop

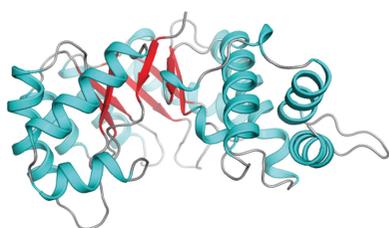
Supporting information: this article has supporting information at journals.iucr.org/d

Methyl transfer between methyltetrahydrofolate and corrinoid molecules is a key reaction in biology that is catalyzed by a number of enzymes in many prokaryotic and eukaryotic organisms. One classic example of such an enzyme is cobalamin-dependent methionine synthase (MS). MS is a large modular protein that utilizes an S_N2 -type mechanism to catalyze the chemically challenging methyl transfer from the tertiary amine (N5) of methyltetrahydrofolate to homocysteine in order to form methionine. Despite over half a century of study, many questions remain about how folate-dependent methyltransferases, and MS in particular, function. Here, the structure of the folate-binding (Fol) domain of MS from *Thermus thermophilus* is reported in the presence and absence of methyltetrahydrofolate. It is found that the methyltetrahydrofolate-binding environment is similar to those of previously described methyltransferases, highlighting the conserved role of this domain in binding, and perhaps activating, the methyltetrahydrofolate substrate. These structural studies further reveal a new distinct and uncharacterized topology in the C-terminal region of MS Fol domains. Furthermore, it is found that in contrast to the canonical TIM-barrel $\beta_8\alpha_8$ fold found in all other folate-binding domains, MS Fol domains exhibit a unique $\beta_8\alpha_7$ fold. It is posited that these structural differences are important for MS function.

1. Introduction

Cobalamin (Cbl)-dependent methionine synthase (MS) is an essential enzyme that is present in many organisms. The role of MS in cells is to catalyze the formation of tetrahydrofolate (H_4 folate) and methionine from methyltetrahydrofolate (CH_3 - H_4 folate) and homocysteine (Hcy) substrates (Fig. 1*a*). Given its central role in metabolism, it is unsurprising that dysfunctions in MS are linked to human disease, including Cbl neuropathy (Metz, 1992), megaloblastic anemia and homocysteinemia (Watkins *et al.*, 2002). In mammals, MS is a monomeric modular protein comprised of four modules arranged as follows (Fig. 1*b*): an N-terminal homocysteine-binding (Hcy) domain, a folate-binding (Fol) domain, a Cbl-binding (Cob) domain and a C-terminal *S*-adenosyl-methionine-binding (AdoMet) domain. As their names imply, each module specifically binds either Hcy, CH_3 - H_4 folate, Cbl or AdoMet.

The MS Cbl cofactor is representative of a class of biological cofactors called corrinoids. Corrinoids are organometallic compounds characterized by a tetrapyrrole ring that harbors a Co-atom center. In MS, the Co atom is found in three different oxidation states: cobalt(I), cobalt(II) and cobalt(III) (Matthews *et al.*, 2008). Cobalt(I) species are among nature's most potent nucleophiles (Schrauzer *et al.*,



1968; Banerjee, 1999) and are generally utilized to perform a variety of ‘difficult’ chemistries including reductive dehalogenations (Payne *et al.*, 2014; Bommer *et al.*, 2014). Corrinoid-dependent enzymes using cobalt(I) promote a wide range of chemistries; however, the most common use of cobalt(I) corrinoids is as cofactors in methyl-transfer reactions (Matthews *et al.*, 2008). For example, in addition to MS, cobalt(I) corrinoid species are used by the methyltransferase subunit (MeTr) of acyl-CoA synthase and CO dehydrogenase enzymes in the CO/CO₂-fixation pathway of anaerobic microbes, and by the MeTr and CFeSP proteins to synthesize methane from acetate in *Moorella thermoacetica* (Ragsdale & Pierce, 2008; Kung *et al.*, 2012).

In contrast to MeTr and CFeP, which are discrete proteins carrying the folate substrates and corrinoid cofactor, all MS enzymes carry these molecules in modules next to each other in the same polypeptide chain. While MS enzymes typically exist as four modules, there are exceptions to this module architecture (Fig. 1*b*). In *Thermotoga maritima*, for example, MS is encoded as a two-component protein, with the Hcy, Fol and Cob modules encoded in one gene and the AdoMet module encoded in a separate open reading frame (Huang *et al.*, 2007; Fig. 1*b*). Despite the differences in subunit/module composition, most of the available structures of Fol modules from these corrinoid methyltransferases appear to exhibit the same classical $\beta_8\alpha_8$ fold (Matthews *et al.*, 2008) of the triose-phosphate isomerase (TIM) barrel (Goldman *et al.*, 2016). The available structures include folate-binding modules from several members of the MeTr (methyltransferase) subfamily [cd00740 in the NCBI Conserved Domain Database (CDD); Marchler-Bauer *et al.*, 2015] of the pterin-binding superfamily (cd00423 in the CCD; PDB entry 2e7f; Doukov *et al.*, 2000) and dihydropterolate synthase [DHPS; cd00739 subfamily; PDB entries 1ajz (Achari *et al.*, 1997) and 1ad4 (Hampele *et al.*, 1997)].

Similar to *Homo sapiens* MS, the thermophilic bacterium *Thermus thermophilus* possesses a four-module Cbl-dependent MS. We selected *T. thermophilus* MS as a model for studying the structural and functional properties of MS because the enhanced stability of thermophiles can facilitate the crystallization of challenging proteins. Here, we describe the structure of the Fol module of *T. thermophilus* MS in the presence and absence of the CH₃-H₄folate substrate. We find that the CH₃-H₄folate-binding environment is similar to those of previously described folate-dependent methyltransferases, highlighting the conserved role of this module in binding and perhaps activating the CH₃-H₄folate substrate. Additionally, our structures reveal a

significant difference in the C-terminal region of most MS Fol modules compared with other folate-binding or pterin-binding proteins. Our studies demonstrate that the structures of most MS Fol modules display a ‘twist’ on the classical $\beta_8\alpha_8$ fold, and are instead best described as having a $\beta_8\alpha_7$ barrel. Therefore, two distinct variations of the TIM-barrel fold are found in pterin-binding proteins.

2. Materials and methods

2.1. Macromolecule production

A gene encoding full-length *T. thermophilus* MS was PCR-amplified from a *T. thermophilus* expression vector (Yokoyama *et al.*, 2000) purchased from RIKEN. The amplified fragment was subsequently cloned into a pMCSG7 plasmid using ligation-independent cloning (Stols *et al.*, 2002) to generate the pMCSG7(tMS^{wt}) plasmid. An expression vector for a His761Lys full-length MS mutant (His761 is found in the Cob module) was generated by site-directed mutagenesis using a QuikChange Mutagenesis Kit (Agilent), and is named pMCSG7(tMS^{H761K}). The coding region of the Fol module (Gln364–His648) was amplified from pMCSG7(tMS^{wt}) and inserted into an empty pMCSG7 plasmid as described above [pMCSG7(tMS^{Fol})]. All of our constructs contain N-terminal hexahistidine-fusion tags (His₆ tags) followed by a cleavable *Tobacco etch virus* (TEV) protease site.

Proteins were expressed in *Escherichia coli* BL21 Star(DE3) cells (Novagen). For expression of the full-length MS protein, the cells were grown to an optical density of 0.6–0.8 in LB medium at 37°C and were then induced by the addition of

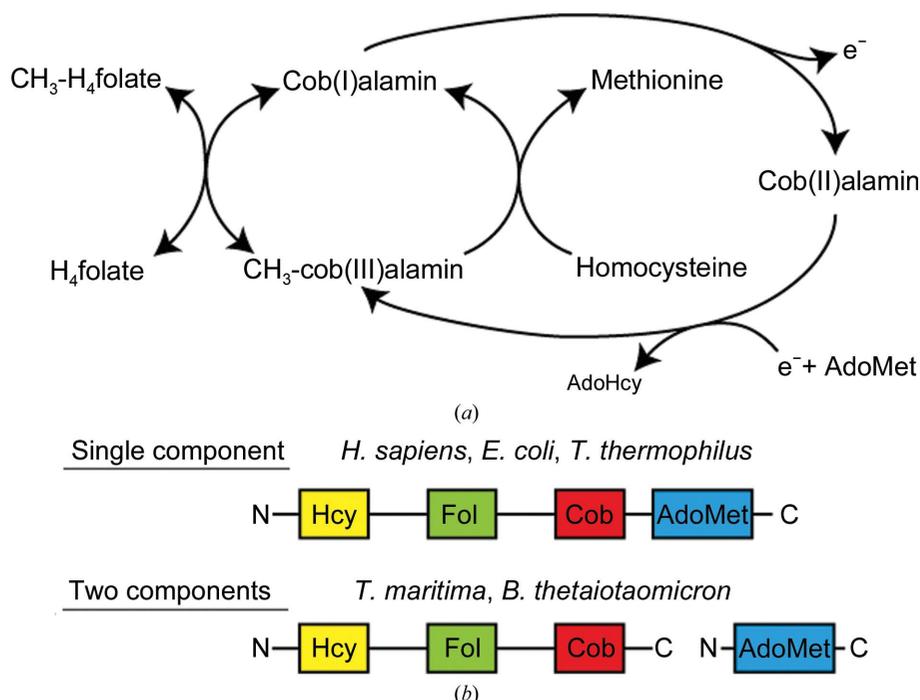


Figure 1
(a) The catalytic cycle of methionine synthase. AdoHcy, adenosylhomocysteine; AdoMet, adenosylmethionine; H₄folate, tetrahydrofolate. (b) The domain structure of cobalamin-dependent methionine synthase.

Table 1
X-ray crystallographic data-collection and refinement statistics.

	MS Fol (1)	MS Fol + CH ₃ -H ₄ folate (2)	MS Fol + CH ₃ -H ₄ folate (3)
Data collection			
Beamline	GM/CA 23-ID-B, APS	GM/CA 23-ID-B, APS	GM/CA 23-ID-B, APS
Wavelength (Å)	1.0332	1.0332	1.0332
Temperature (K)	100	100	100
Resolution (Å)	50.0–2.10 (2.18–2.10)	50.0–2.10 (2.16–2.10)	50.0–2.40 (2.54–2.40)
Crystal-to-detector distance (mm)	300	300	350
Rotation range per image (°)	0.5	0.2	0.2
Total rotation range (°)	130	110	40
Exposure time per image (s)	1	0.2	0.2
Space group	<i>P</i> ₄ ₃ ₂ ₁ ²	<i>P</i> ₆ ₁ ₂₂	<i>P</i> ₆ ₁ ₂₂
Unit-cell parameters (Å, °)	<i>a</i> = 55.2, <i>b</i> = 55.2, <i>c</i> = 197.1, $\alpha = \beta = \gamma = 90$	<i>a</i> = 111.5, <i>b</i> = 111.5, <i>c</i> = 242.7, $\alpha = \beta = 90, \gamma = 120$	<i>a</i> = 188.5, <i>b</i> = 188.5, <i>c</i> = 247.1, $\alpha = \beta = 90, \gamma = 120$
Observed reflections	123053	561666	453517
Unique reflections	18533	52825	187437
<i>R</i> _{meas} (%)	5.9 (54.0)	9.5 (79.1)	10.3 (73.1)
<i>R</i> _{merge} (%)	5.1 (41.6)	8.7 (71.6)	8.1 (56.3)
<i>R</i> _{p.i.m.} (%)	2.9 (33.4)	3.8 (32.9)	6.1 (45.6)
<i>I</i> (σ (<i>I</i>))	17.0 (2.1)	11.3 (1.9)	11.3 (2.03)
CC _{1/2}	0.999 (0.783)	0.999 (0.909)	0.996 (0.692)
Multiplicity	5.8 (3.7)	10.6 (10.2)	4.5 (4.4)
Completeness (%)	98.0 (93.6)	99.9 (99.9)	99.4 (98.4)
Overall <i>B</i> (Wilson plot) (Å ²)	32.6	42.8	39.6
Refinement			
Resolution range	53.14–2.10	96.55–2.10	163.2–2.40
No. of reflections			
Work set	17499	50064	95477
Test set	955	2667	5056
<i>R</i> _{work} / <i>R</i> _{free} (%)	20.2/25.9	19.4/22.2	19.0/23.0
No. of non-H atoms			
Protein	2217	4434	13273
Water	67	128	322
C2F [†]	0	66	198
<i>B</i> factors (Å ²)			
Protein	50.4	61.7	50.5
Water	52.0	57.1	43.1
C2F [†]	n/a	63.5	35.3
R.m.s. deviations			
Bond lengths (Å)	0.008	0.009	0.009
Bond angles (°)	1.300	1.420	1.363
Ramachandran plot (%)			
Favored	96.8	97.4	97.0
Allowed	2.5	2.4	2.9
Outliers	0.7	0.2	0.1
<i>MolProbity</i> score	1.10 [100th percentile]	1.00 [100th percentile]	1.18 [100th percentile]
PDB code	5von	5vop	5voo

[†] 6S-5-Methyltetrahydrofolate.

isopropyl β-D-1-thiogalactopyranoside (final concentration of 0.1 mM; Promega) followed by overnight growth at 25°C. For expression of the single Fol module, *E. coli* BL21 Star(DE3)/pMCSG7(tMS^{Fol}) cells were inoculated in auto-induction medium (Studier, 2005) and cultured at 30°C for 22–24 h. The cells were collected by centrifugation and stored at –80°C.

Purification of full-length *T. thermophilus* MS without Cbl (apo *T. thermophilus* MS^{H761K}) was performed as described below. *E. coli* BL21 Star(DE3)/pMCSG7(tMS^{H761K}) cells were resuspended in buffer [50 mM potassium phosphate buffer (KPB) pH 7.2, 1 mM phenylmethylsulfonyl fluoride, 0.5 μg ml^{–1} lysozyme], lysed by sonication and pelleted by centrifugation (10 000g for 20 min at 4°C). The resulting supernatant was subjected to a heat step by incubation at 70°C for 20 min. The heat-treated suspension was subsequently

centrifuged (10 000g for 20 min) and filtered using a Millex Filter Unit (0.45 μm; Millipore, Darmstadt, Germany). The filtrate was loaded onto a 5 ml nickel-affinity column (HiTrap Chelating; GE Healthcare Life Sciences) at room temperature (in buffer *A*: 50 mM KPB pH 7.2, 50 mM imidazole). The column was washed with at least five column volumes of buffer *B* (buffer *A* + 0.3 M sodium chloride). The bound protein was eluted with ~4 column volumes of 50 mM KPB pH 7.2, 125 mM imidazole. Eluates containing His₆-tagged protein were pooled together and were dialyzed overnight at 4°C against 50 mM KPB pH 7.2. The resulting dialyzed protein was concentrated and stored at 4°C.

To express and purify the Fol module of *T. thermophilus* MS, pMCSG7(tMS^{Fol}) was transformed into *E. coli* BL21 Star(DE3) cells. Cell lysis followed by heat treatment was performed as described above for the full-length MS preparation. After adding imidazole to a final concentration of 20 mM, the crude protein solution was loaded onto a nickel-affinity column. The column was washed with buffer *B* and the bound protein was eluted with buffer *C* (buffer *B* + 200 mM imidazole). The eluate was dialyzed overnight against 50 mM Tris–HCl pH 7.4 in the presence of His₆-tagged TEV protease at 4°C overnight, after which the

dialysate was incubated at 70°C for 15 min to remove TEV protease. Purified MS Fol module was concentrated and stored at 4°C.

2.2. Crystallization

For crystallization purposes, full-length apo *T. thermophilus* MS^{H761K} was run through a size-exclusion column (S200; GE Healthcare) equilibrated with 25 mM Tris–HCl pH 7.5, 0.1 M NaCl. The full-length *T. thermophilus* MS and Fol module samples were concentrated to ~15 and ~25 mg ml^{–1}, respectively. All protein samples were filtered through a Spin-X centrifuge filter (Corning Costar) prior to crystallization screening.

Crystals of full-length *T. thermophilus* MS were obtained by the sitting-drop vapor-diffusion method by mixing 0.3 μl

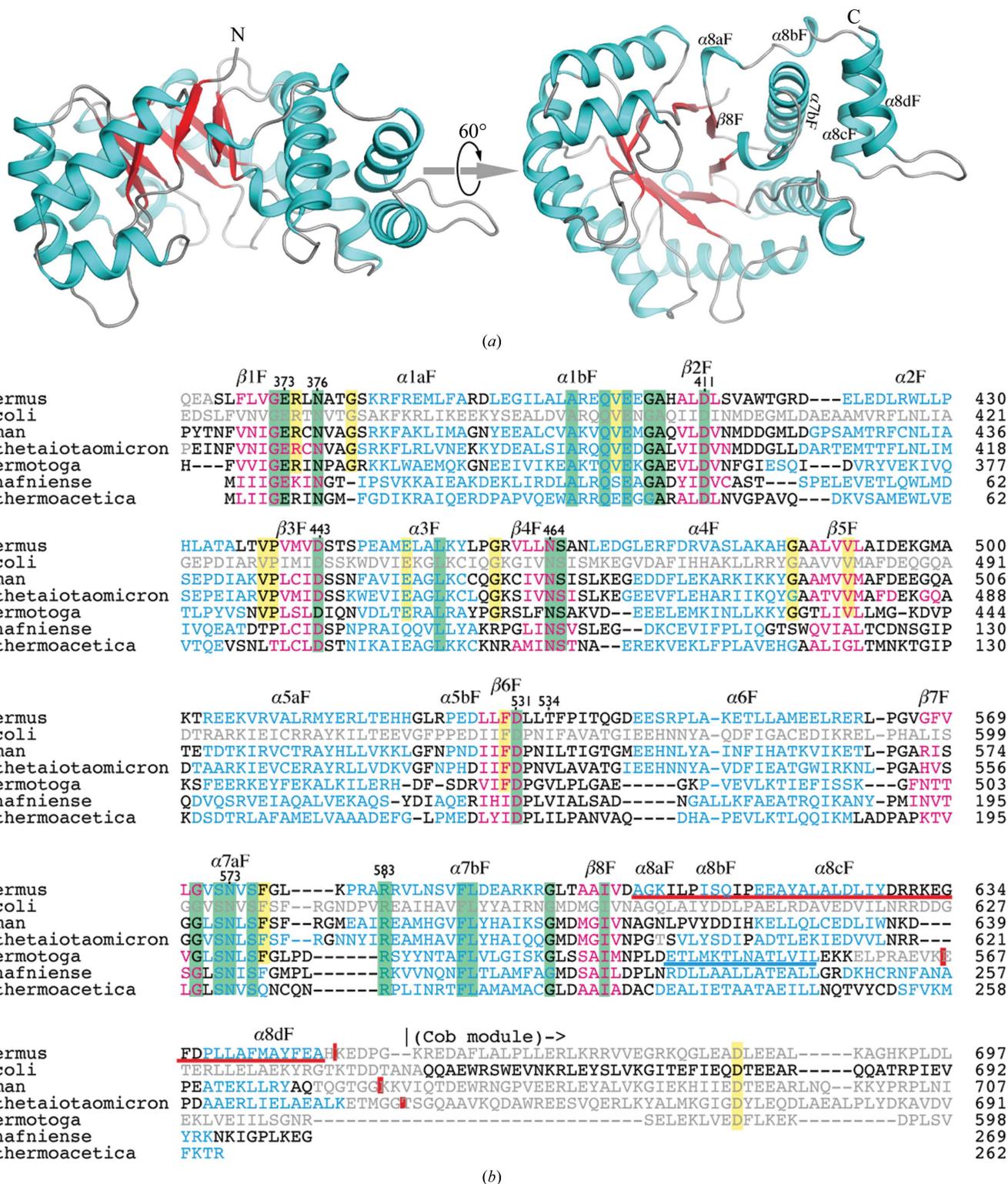


Figure 2 Structure of the *T. thermophilus* MS folate-binding module. (a) The crystal structure of the *T. thermophilus* MS Fol module is shown in cartoon mode. Helices and sheets are colored cyan and red, respectively. (b) Amino-acid alignment of the folate-binding modules of MS and MeTr. Thermus, *Thermus thermophilus* MS (GenBank accession No. NC_006461); E.coli, *Escherichia coli* MS (J04975); human, *H. sapiens* MS (U73338); B.thetaiotaomicron, *Bacteroides thetaiotaomicron* VPI-5482 MS (NC_004663); Thermotoga, *Thermotoga maritima* MS (NC_000853); D.hafniense, *Desulfitobacterium hafniense* DCB-2 methyltetrahydrofolate-corrinoid/iron-sulfur protein methyltransferase (Dhaf_0722; ACL18786); M.thermoacetica, *Moorella thermoacetica* 5-methyltetrahydrofolate-corrinoid/iron-sulfur protein co-methyltransferase (AcsE; Q46389). Protein portions forming helices and sheets are colored cyan and magenta, respectively. Amino-acid chains that comprise structurally undetermined regions are shown in gray. Residues conserved only in MS are highlighted in yellow and those conserved across all methyltransferases are highlighted in green. The last helix of the Fol module in *T. thermotoga* MS (α 8F) is underlined in blue and the C-terminal region including helices α 8aF, α 8bF, α 8cF and α 8dF of *T. thermophilus* MS is underlined in red. The red vertical lines indicate the C-terminal ends of the expressed and crystallized protein constructs.

protein solution (14 mg ml^{-1}) with $0.3 \mu\text{l}$ reservoir solution. Full-length apo *T. thermophilus* MS was mixed with 0.8 M succinic acid pH 7.0 at 20°C and afforded crystals of $\text{CH}_3\text{-H}_4\text{folate-free MS Fol module (1)}$. Crystals of **1** were cryoprotected for 5 min before being flash-cooled in liquid nitrogen by transfer to a solution consisting of 20% glycerol, 0.8 M succinic acid pH 7.0.

For the crystallization of single Fol module MS constructs, the hanging-drop vapor-diffusion method was used. The Fol module of *T. thermophilus* MS (15 mg ml^{-1}) was mixed with reservoir solution (0.1 M sodium citrate pH 4.5, 18% PEG 3350, 0.2 M trisodium citrate) in a 1:1 ratio at 4°C . Crystals of the Fol module with native $\text{CH}_3\text{-H}_4\text{folate}$ present (**2**) were cryoprotected for 5 min before being flash-cooled in liquid nitrogen by transfer to a solution consisting of 0.1 M sodium citrate pH 4.5, 18% PEG 3350, 0.2 M trisodium citrate, 20% glycerol.

Crystals of the Fol module grown in the presence of externally provided $\text{CH}_3\text{-H}_4\text{folate}$ (**3**) were transferred to a solution consisting of 0.1 M sodium citrate pH 4.5, 22% PEG 3350, 0.2 M trisodium citrate, 20% glycerol, $\sim 1.5 \text{ mM}$ (6*S*)-5- $\text{CH}_3\text{-H}_4\text{folate}$ (Eprova), in which they were stored overnight before they being flash-cooled in liquid nitrogen.

2.3. Data collection and processing

Diffraction data for all crystals were collected at 100 K on the GM/CAT-CAT 23-ID-B beamline at Advanced Photon Source (APS), Argonne National Laboratory using a MAR300 detector for **1** and using a DECTRIS EIGER X 16M detector for **2** and **3**. All data sets were processed with XDS (Kabsch, 2010): to 2.1 \AA resolution for **1**, 2.4 \AA resolution for **2** and 2.1 \AA resolution for **3**. Data-collection and processing statistics are summarized in Table 1. The crystals of **1** belonged to

space group $P4_32_12$ (unit-cell parameters $a = b = 55.2$, $c = 197.1 \text{ \AA}$) with one molecule in the asymmetric unit (Matthews coefficient $V_M = 2.3 \text{ \AA}^3 \text{ Da}^{-1}$, 45.9% solvent content) and the crystals of **3** belonged to space group $P6_122$ (unit-cell parameters $a = b = 188.5$, $c = 247.1 \text{ \AA}$) with six molecules in the asymmetric unit (Matthews coefficient

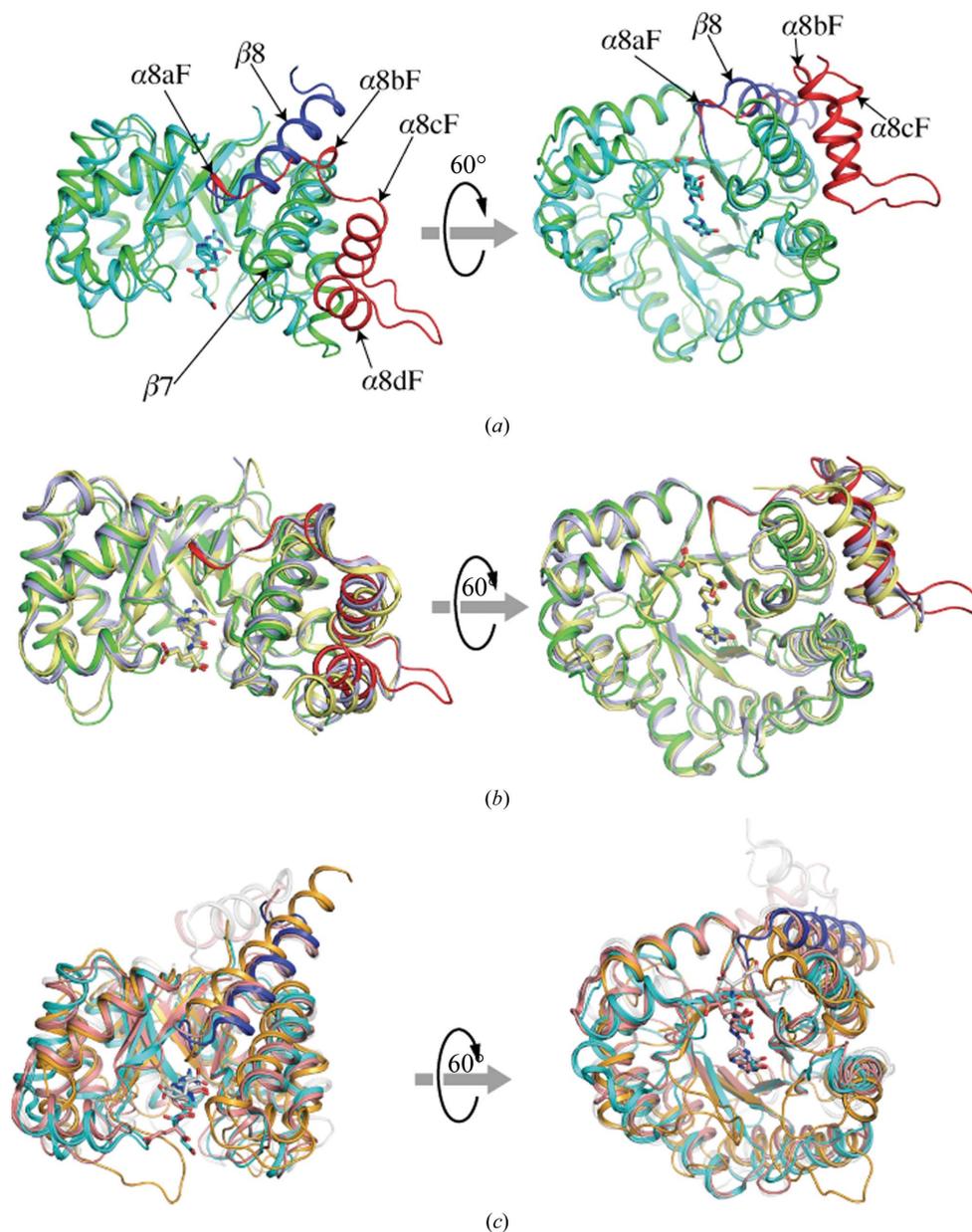


Figure 3

Demonstration of the two topology types observed in the C-terminal region of folate-binding modules. (a) Superposition of the Fol modules from *T. thermophilus* and *T. thermotoga* MS. The Fol module of *T. thermophilus* MS is illustrated in cartoon mode and shown in green, except for the C-terminal four helices, which are shown in red. *T. thermotoga* MS (PDB entry 1q8j) is shown in cyan, except for the last helix ($\alpha 8\text{F}$), which is colored blue. (b) Structural alignment of the Fol modules of MS. The Fol modules of *T. thermophilus* MS [colored green and red as in (a)], *H. sapiens* (pale blue; PDB entry 4ccz) and *B. thetaiotaomicron* (yellow; PDB entry 3k13) are drawn in cartoon mode. (c) Structural alignment of pterin-binding superfamily members. The Fol module of *T. thermotoga* MS is shown in cyan and blue as in (a). The folate-binding domains of MeTr from *M. thermoacetica* (PDB entry 2e7f) and MT2DH from *D. hafniense* (PDB entry 4o1e) are displayed in pink and gray, respectively, and the last nonconserved helices are transparent. A monomer of DHPS from *S. aureus* (PDB entry 1ad4) is colored orange.

$V_M = 3.2 \text{ \AA}^3 \text{ Da}^{-1}$, 61.5% solvent content), whereas the crystals of **2** belonged to space group $P6_122$ (unit-cell parameters $a = b = 111.5$, $c = 242.7 \text{ \AA}$) with two molecules in the asymmetric unit (Matthews coefficient $V_M = 3.5 \text{ \AA}^3 \text{ Da}^{-1}$, 64.5% solvent content).

The crystals of **1** were obtained using full-length *T. thermophilus* MS; however, the unit cell was far too small to correspond to the full-length polypeptide based on the Matthews coefficient. This was confirmed by SDS–PAGE analysis of the crystals, which showed a molecular weight of ~ 31 kDa which corresponds to a single module. Therefore, the crystals corresponding to a single module were formed after proteolytic cleavage, possibly by trace contamination with protease during crystallization. Since we had not ascertained the nature of this single module, we used all four MS modules as search models for a molecular-replacement solution. Initial phases were obtained with *Phaser* (McCoy *et al.*, 2007) only when the Fol module of *T. maritima* MS (PDB entry 3bof; Koutmos *et al.*, 2008) was used as a search model. Initial phases for **2** and **3** were obtained using the resulting structure of **1** as a search model.

2.4. Structure solution and refinement

All three structures were initially refined with *PHENIX* (Adams *et al.*, 2002). Refinement included simulated annealing in torsional and Cartesian space, coordination minimization and restrained individual B -factor adjustment with maximum-likelihood targets. *REFMAC5* (Murshudov *et al.*, 2011) was then employed for restrained refinement using isotropic individual B factors with maximum-likelihood targets. In all three refinements the Babinet model for bulk-solvent scaling was utilized. Refinement was followed by model building and modification with *Coot* (Emsley *et al.*, 2010). Several iterative rounds of refinement followed by model building and modification were performed. In the early rounds of refinement, no ligands were fitted for structures **2** and **3**.

After the protein model was complete and water molecules had been assigned, clear electron density was fitted in the active sites of **2** and **3** and subsequently refined as $\text{CH}_3\text{-H}_4\text{folate}$ (C2F) ligands. For **2**, since no external folate was provided the electron density corresponds to native folate substrate. We considered both the $\text{CH}_3\text{-H}_4\text{folate}$ (C2F) substrate and the H_4folate (THG) product as potential ligands that could occupy the electron density in the active site. When H_4folate was modeled there was strong residual electron density near the N5 position, clearly corresponding to the expected position of the methyl group in $\text{CH}_3\text{-H}_4\text{folate}$. $\text{CH}_3\text{-H}_4\text{folate}$ fits well in the composite OMIT electron density, as shown in Supplementary Fig. S1. Therefore, we concluded that the native ligand corresponds to $\text{CH}_3\text{-H}_4\text{folate}$. It must be noted that there is residual but weak electron density extending from the glutamate tail, indicating that the native ligand has a polyglutamate moiety. Owing to the poor electron density in this region we did not attempt to model such a polyglutamate tail, and only a single glutamate is present in our model.

The final model was used to create an unbiased refined composite $2F_o - F_c$ OMIT map with *PHENIX* (Supplementary Fig. S1). Crystallographic information as well as refinement statistics are provided in Table 1. The geometric quality of the model and its agreement with the structure factors were assessed with *MolProbity* (Chen *et al.*, 2010). For **1**, *MolProbity* reported a clashscore and a *MolProbity* score of 1.11 (100th percentile) and 1.10 (100th percentile) for **1**, while 96.8% of the residues were in the favored Ramachandran plot regions, with 0.7% residues in outlier regions. For **2**, the reported values were 1.00 (100th percentile) and 1.97 (100th percentile), respectively, while 97.9% of the residues were in the favored Ramachandran plot regions, with 0.4% residues in outlier regions. For **3**, the reported values were 1.18 (100th percentile) and 2.24 (100th percentile), respectively, while 97.0% of the residues were in the favored Ramachandran plot regions, with 0.1% of residues in outlier regions. Figures displaying crystal structures were generated with *PyMOL* (v.1.8; Schrödinger; <http://www.pymol.org>).

3. Results

3.1. Crystal structure of the Fol module of *T. thermophilus* MS

The structure of the *T. thermophilus* MS Fol module is shown in Fig. 2(a). The overall structure is in good agreement with previously published structures of folate-binding modules from other MeTr family members. We aligned the sequences of selected MeTr members, including the Fol module of *T. thermophilus* MS, as shown in Fig. 2(b), to highlight invariant and highly conserved residues among various folate-binding modules. To assess the potential differences between MS Fol modules from different species, we aligned our structure of the MS Fol module from *T. thermophilus* with that of the previously described MS module from *T. maritima* (PDB entry 1q8j; Evans *et al.*, 2004; Fig. 3a). The Fol modules from *T. thermophilus* and *T. maritima* MS share 25% identity (based on 292 amino acids). Moreover, their structural superposition was indicative of the structural similarity between these two modules, with an r.m.s.d. for C^α atoms of 1.98 \AA (222 residues).

In addition to the anticipated similarities, we also observed a region with significant differences. While the eight β -strands in the center of the TIM barrel and the first seven α -helices that flank the central barrel are comparable between the *T. maritima* and *T. thermophilus* MS Fol modules, the region corresponding to the eighth α -helix at the C-terminus of the TIM barrels is different in *T. thermophilus* MS. In typical TIM barrels, such as the *T. maritima* MS Fol module (Figs. 2b and 3a), the eighth β -strand is followed by a short four-residue loop (538–541) and the eighth α -helix (14 residues in length). This creates a characteristic antiparallel strand–turn–helix motif that places the eighth helix almost parallel to the barrel. In this case, the eighth α -helix forms an α/α interface and interacts with both the surrounding $\alpha 1$ and $\alpha 7$ helices (Fig. 3a). In contrast to this, the Fol module of *T. thermophilus* MS does not have the antiparallel helix–turn–helix motif present at the

C-terminus. Instead, we observe a unique arrangement that is distinct from other folate-binding modules.

In *T. thermophilus* MS Fol the eighth β -strand is followed by an 11-residue loop containing two 3_{10} -helices ($\alpha 8aF$ and $\alpha 8bF$) that extend away from the barrel and bend to wrap around the solvent-exposed side of the $\alpha 7$ helix. This loop is followed by two antiparallel helices, $\alpha 8cF$ and $\alpha 8dF$, that stack against the outer sides of barrel helices $\alpha 6$ and $\alpha 7$ (Figs. 2*a* and 3*a*). The resulting arrangement resembles a helical bundle that displays an extensive interface with multiple electrostatic and hydrophobic interactions and a total buried area of 1000 \AA^2 as calculated by the PISA server (Krissinel & Henrick, 2007). Consequently, whereas the *T. maritima* MS Fol module displays a typical and canonical $\beta_8\alpha_8$ barrel, the Fol module of *T. thermophilus* MS is best described as a $\beta_8\alpha_7$ barrel. Thus, the *T. thermophilus* MS Fol module represents the first described example of a noncanonical subclass of TIM barrels.

A DALI server search (Holm & Rosenström, 2010) using the Fol module of *T. thermophilus* MS as a search model identified the Fol module from *H. sapiens* MS (PDB entry 4ccz; Structural Genomics Consortium, unpublished work) as the closest structural match (Z -score = 37.1, r.m.s.d. = 3.6 \AA). The second closest structural homolog was that of the MS Fol module from *Bacteroides thetaiotaomicron* (PDB entry 3k13; Midwest Center for Structural Genomics, unpublished work; Z -score = 27.2, r.m.s.d. = 2.0 \AA). Both of these structures have been deposited in the PDB, but are not discussed in published manuscripts. A structural superposition of the Fol modules of *T. thermophilus*, *H. sapiens* and *B. thetaiotaomicron* MS is graphically illustrated in Fig. 3(*b*). It is evident that these three Fol modules share the same topology, including an identical $\beta_8\alpha_7$ TIM-barrel. These differ from the Fol module of *T. maritima* MS, which exhibits the more common $\beta_8\alpha_8$ fold.

3.2. Comparison of the Fol module of *T. thermophilus* MS with folate-binding and pterin-binding modules from various species

In addition to the findings discussed above, a DALI search for MS Fol modules identified other distinct structural homologs. Most of these also belong to the pterin-binding superfamily (MeTr subfamily; cd00423 in the CDD). For the purposes of comparison, the following three structurally related proteins were selected: the MeTr subunit of acyl-CoA synthase/CO dehydrogenase from *Moorella thermoacetica* (PDB entry 2e7f; Doukov *et al.*, 2007), the methyltransferase component, the MT2DH product of *Dhaf0722*, from *Desulfitobacterium hafniense* DCB-2 (PDB entry 4o1e; Sjuts *et al.*, 2015) and DHPS from *Staphylococcus aureus* (PDB entry 1ad4; Hampele *et al.*, 1997). As shown in Fig. 3(*c*), all three of

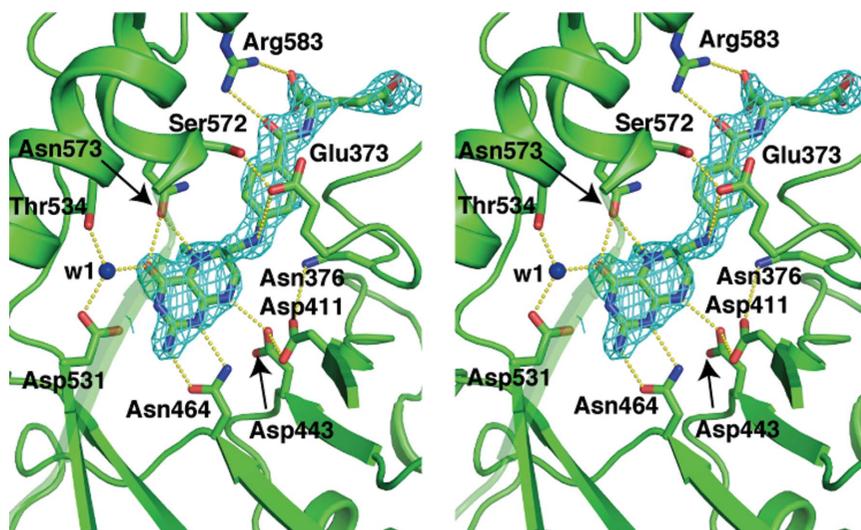


Figure 4

Stereoview of the $\text{CH}_3\text{-H}_4$ folate-binding site. The Fol module is rendered in cartoon mode, while the active-site side chains and the folate substrate are shown as sticks. The weighted and refined $2mF_o - DF_c$ map (blue) shown for the folate substrate is contoured at 1.5σ .

these protein structures share the hallmark $\beta_8\alpha_8$ -barrel motif found ubiquitously throughout this family. The C-terminal topology of these three structures is identical to that of the Fol module of *T. maritima* MS, albeit distinct from that of the Fol module of all other MS enzymes.

3.3. Comparison of folate-binding modes

To determine how the *T. thermophilus* MS Fol module binds its folate substrate, the isolated Fol module was expressed, purified and crystallized in the presence and absence of $\text{CH}_3\text{-H}_4$ folate. We observed electron density corresponding to $\text{CH}_3\text{-H}_4$ folate even in crystals that had not been soaked or co-crystallized with $\text{CH}_3\text{-H}_4$ folate (Fig. 4), suggesting that the Fol module co-purifies in complex with its substrate. Additionally, we solved the structure of a *T. thermophilus* MS Fol module that was proteolytically derived from the full-length MS. In contrast to the isolated module, the crystallized proteolytic fragment did not show any electron density associated with $\text{CH}_3\text{-H}_4$ folate. This allowed us to assess the potential impact on the Fol module structure when $\text{CH}_3\text{-H}_4$ folate binds. Comparison of the substrate-free structure with the substrate-bound structure (Supplementary Fig. S2) showed no significant differences, with r.m.s.d.s of 1.293 and 0.637 \AA based on all atoms (2217) and all backbone atoms (1128), respectively. The folate-bound structure is slightly more open relative to the folate-free structure. When the $\text{CH}_3\text{-H}_4$ folate ligand is present it presses against helix $\alpha 1aF$ and the loop connecting it to $\beta 1F$. This causes helix $\alpha 1aF$, the loop connecting it to $\beta 1F$ and helix $\alpha 1bF$ to move away from the ligand. This is the only ligand-induced change that results in the observed slight opening of the barrel. There is little change in the residues that line the substrate-binding pocket (Supplementary Fig. S2*b*); the majority of these residues are in the same position in both

structures, reflecting a lock-and-key type of fit between the ligand and its binding module.

In *T. thermophilus* MS the pterin ring of folate is recognized through residues that are conserved among all Fol modules. The amino acids comprising the folate-binding site are displayed as sticks in Fig. 5 (Asp411, Asp443, Asn464, Asp531, Asn573 and Thr534); these are all conserved except for Thr534 (Fig. 2*b*). The binding-pocket residues form an extensive hydrogen-bonding network that is nearly identical in all Fol modules. The only differences that are observed in the folate-binding pocket between the various Fol module structures are found near the N3 and O4 folate atoms. For example, in the structure of *T. thermophilus* MS a single water molecule (W1) is present near folate atoms N3 and O4, while two water molecules (W1 and W2) are observed in the other Fol modules (Fig. 5). W1 is present in all available folate-bound structures and mediates a contact between an invariant Asp carboxylate (Asp531 in *T. thermophilus* MS) and the carbonyl of the pterin ring (O4). In the three structures containing a second water

molecule (W2; Figs. 5*b*, 5*c* and 5*d*), W2 forms interactions with the carbonyl O atom of the peptide backbone (Val or Ile), the amino group of an invariant Asn and W1. In contrast, Thr534 of *T. thermophilus* MS interacts directly with W1 (Fig. 5*a*).

Additionally, there are differences in a conserved Ser residue near the *p*-aminobenzoate moiety (Ser572 in *T. thermophilus* MS). The equivalent Ser198 in *D. hafniense* MT2DH (Fig. 5*d*) is proposed to play an important role in substrate activation based on the recent structures of MT2DH-CH₃-H₄folate complexes (Sjuts *et al.*, 2015). In all other Fol modules this Ser residue forms a hydrogen bond to an invariant Glu (Glu373 in *T. thermophilus* MS; Figs. 4 and 6), but in *D. hafniense* MT2DH it drastically relocates and forms a hydrogen bond to a water molecule (W3) that interacts with the pterin N5 and O4 atoms. This water is not observed in other protein-CH₃-H₄folate binary structures (W3 in Fig. 5*d*). The unique Ser arrangement in *D. hafniense* MT2DH may be important for the distinct functionality of MT2DH, which supports the reverse reaction compared with other

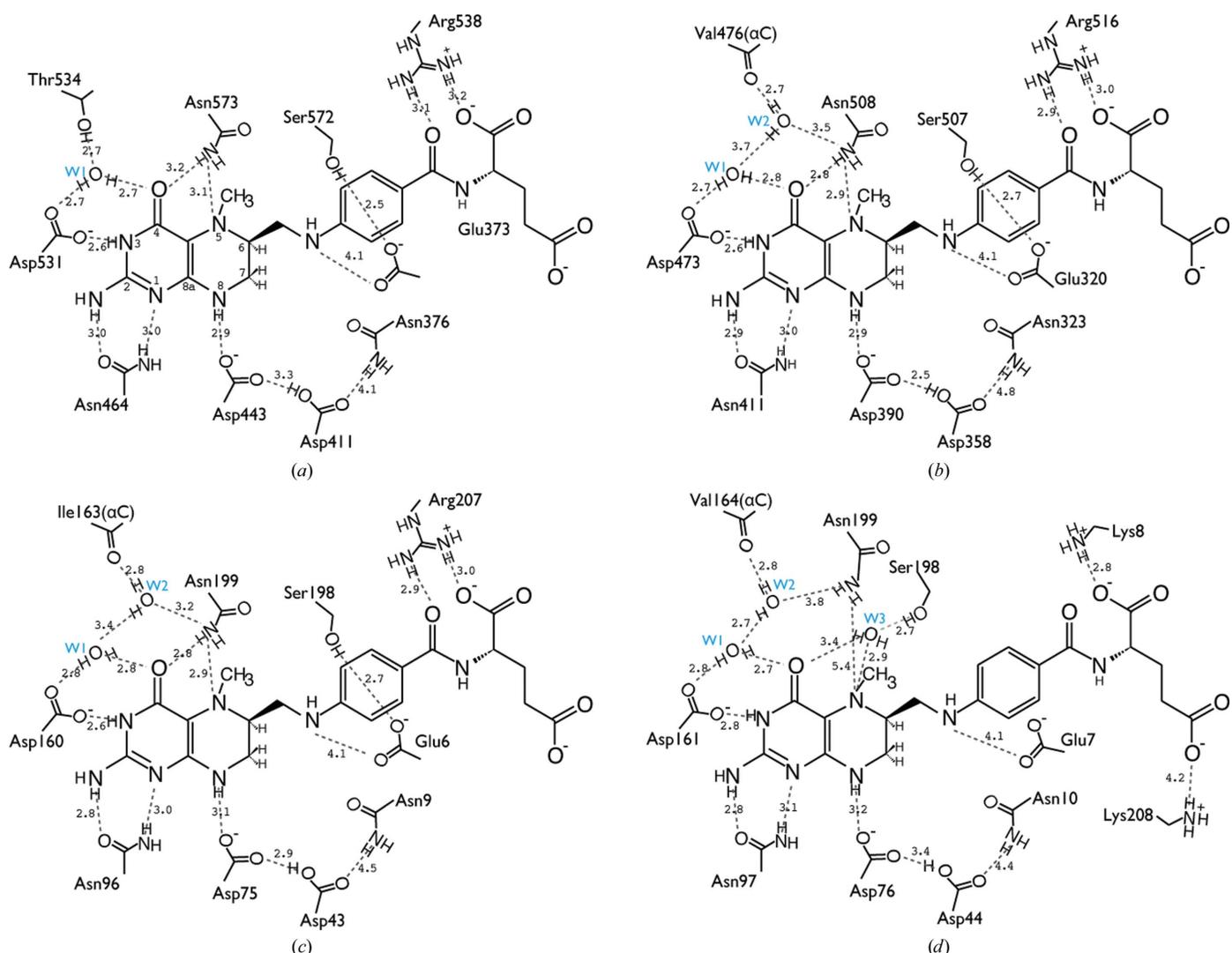


Figure 5
Schematic illustration of the interaction between amino-acid residues and CH₃-H₄folate in *T. thermophilus* MS (a), *T. thermotoga* MS (b), *M. thermoacetica* MeTr (c) and *D. hafniense* MT2DH (d).

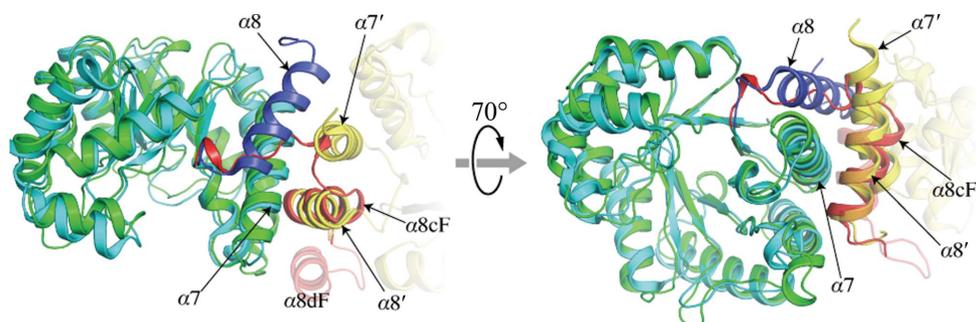


Figure 6

Comparison of the C-terminal regions of the *T. thermophilus* Fol module monomer and the *T. thermotoga* Fol module dimer. The Fol module of *T. thermophilus* MS (green and red) is superimposed on one of the *T. thermotoga* MS Fol modules (cyan and blue) as shown in Fig. 3(a). The adjacent Fol module of the *T. thermotoga* MS dimer is partially drawn and shown in yellow.

folate-dependent methyltransferases: methylation of the folate substrate (Sjuts *et al.*, 2015). In the other available structures of Fol modules with a substrate bound, including those of MS enzymes, instead of W3 an invariant Asn [Asn199 in MeTr (Fig. 5c), and Asn573 and Asn508 in *T. thermophilus* MS and *H. sapiens* MS, respectively (Figs. 5a and 5b)] directly interacts with the O4 and N5 atoms of the folate substrate. It has been proposed that this invariant Asn is essential for catalytic activity, but not for substrate binding, in all folate-dependent methyltransferases (Doukov *et al.*, 2007). Lastly, three invariant Asp residues (Asp411, Asp443 and Asp531) are expected to be necessary for substrate activation based on studies of *E. coli* MS (Smith & Matthews, 2000).

4. Discussion

Our studies of the Fol module of *T. thermophilus* MS revealed a unique variation of the prototypical $\beta_8\alpha_8$ TIM-barrel fold that is present in all other pterin-binding proteins (Marchler-Bauer *et al.*, 2015). The pterin-binding superfamily (cd00423) is divided into two subfamilies: the DHPS (cd00739) and MeTr (cd00740) subfamilies. Based on our results, we propose that the MeTr subfamily should be further divided into two subgroups: one exhibiting the classical $\beta_8\alpha_8$ fold and a second displaying the distinct $\beta_8\alpha_7$ fold described here and only found thus far in the Fol modules of Cbl-dependent methionine synthases.

Prior to this work, the only previously published structure of a Fol module was from *T. maritima*. Here, we compare the Fol domain of MS from *T. maritima* with our structure from *T. thermophilus*, and also with structures deposited in the PDB (but not published) from *H. sapiens* and *B. thetaiotaomicron*. We find that with the exception of *T. maritima* MS, all MS Fol modules appear to be members of the new MeTr subfamily containing a distinct $\beta_8\alpha_7$ TIM barrel. We first speculated that this may be because in contrast to the majority of species, which encode their MS proteins in a single gene, *T. maritima* MS encodes the first three modules (Hcy/Fol/Cob) in a single gene and the last activation module (AdoMet) separately (Huang *et al.*, 2007). However, this seems unlikely because *B. thetaiotaomicron* MS, like *T. maritima* MS, is also encoded

in a two-component system (NCBI locus tags BT_0180 for the Hcy, Fol and Cob modules and BT_0249 for the putative AdoMet module) yet its Fol module still exhibits a $\beta_8\alpha_7$ -barrel fold. Another explanation for this difference may arise from the observation that among the crystallized MS Fol modules only *T. maritima* MS Fol is involved in homodimer formation both *in crystallo* and in solution. Close examination of the *T. maritima* MS dimerization interface (Fig. 6) reveals that it is formed mainly

through interactions of helices α_7 and α_8 with the same helices (α_7' and α_8') from an adjacent molecule. The total buried surface area in the interface is $\sim 1044 \text{ \AA}^2$, which is nearly identical to that found in the equivalent C-terminal helical bundle present in the other MS Fol modules. Interestingly, superposition of the Fol modules of *T. thermophilus* and *T. maritima* MS revealed that in *T. maritima* the MS Fol homodimer interface helix α_7' occupies the exact same position as helix α_8cF in the *T. thermophilus* MS Fol module (Fig. 6). Furthermore, *T. maritima* MS was crystallized with a truncated C-terminus lacking residues in the linker between the Fol and Cob modules (Fig. 2b). The truncated *T. maritima* construct lacked amino acids corresponding to the entire helix α_8dF , in contrast to the three MS Fol modules with a $\beta_8\alpha_7$ fold. It is therefore possible that the classical $\beta_8\alpha_8$ fold found in the *T. maritima* MS Fol module may be an artifact of the crystallization construct owing to the absence of the linker region residues that are present in all of the other determined MS folate structures. This would allow the truncated *T. maritima* MS Fol module to fold differently, inhibiting the ability of this C-terminal region to fold into the noncanonical TIM-barrel ($\beta_8\alpha_7$) structure seen in the structures of MSs from all other sources. However, an alternative justification for this variation may lie in the differences found both in the C-terminal region of the Fol module and in the linker region connecting the Fol module to the subsequent Cob module. In *T. maritima* MS there are 29 amino acids between the eighth strand of the folate domain (β_8F) and the first helix of the N-terminal four-helical bundle of the Cob module. In the other MSs the linker regions are substantially longer, ranging between 59 and 63 residues. Given the differences in the linker region in *T. maritima* MS, we believe that the latter is most likely to represent an outlier and has a distinct method of connecting the Fol module to the Cob module. Moreover, close examination of the C-terminal region, and especially helices α_7 and β_8 of the Fol modules, reveal significant variations in the amino-acid sequences that might explain the different conformations. Specifically, we find four major differences between Fol modules with a classical fold (e.g. *T. maritima*) and a 'twist' fold (e.g. *T. thermophilus*) (Fig. 7 and Supplementary Fig. S3).

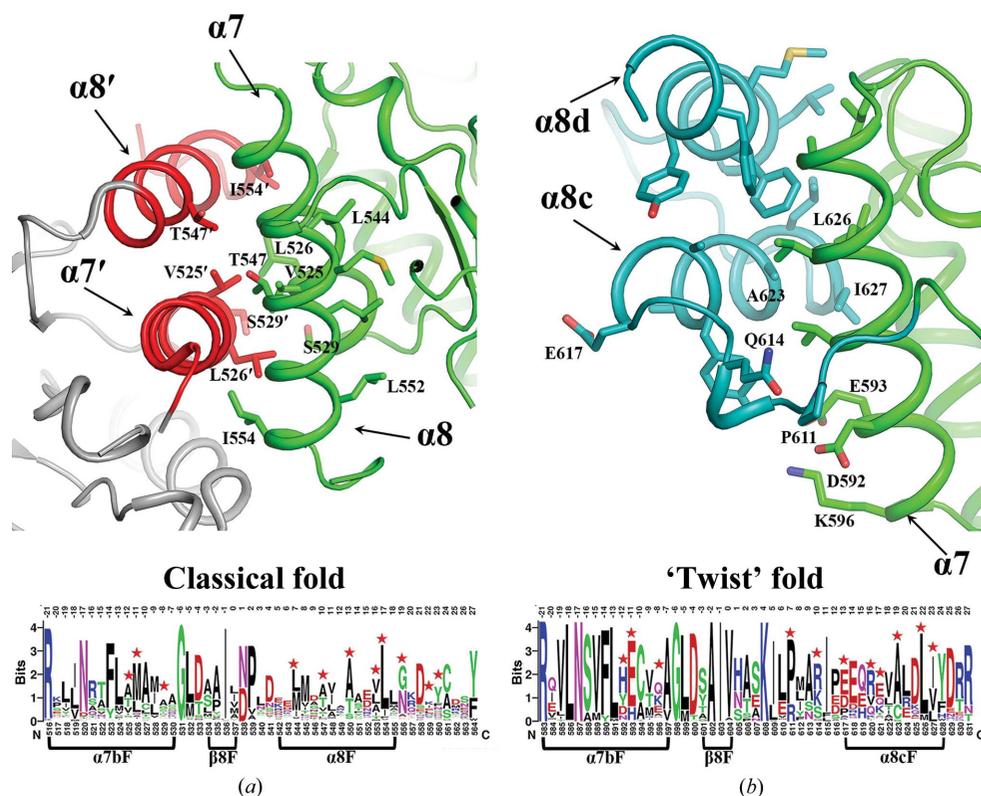


Figure 7

The top panels show close-ups of the two distinct interfaces formed in Fol modules with a $\beta_8\alpha_8$ and a $\beta_8\alpha_7$ fold in cartoon mode. Important residues that form part of the interfaces are shown as sticks. The bottom panels display the graphical sequence variance between the C-terminal regions in the classical ($\beta_8\alpha_8$) and 'twist' ($\beta_8\alpha_7$) folds, as generated with *WebLogo* (Crooks *et al.*, 2004). More than 300 sequences (with representative sequences exhibiting between 90 and 15% sequence identity) for each of the different folds were used in *WebLogo*. The numbering at the bottom corresponds to the *T. maritima* (left) and *T. thermophilus* (right) sequences, whereas the numbering at the top is used as a reference and corresponds to the positions upstream (negative) or downstream (positive) of the last residue in the β_8 strand. In (a) helix α_8 (green) is involved in intermolecular interactions with helices α_7 and α_1 (green) and mainly hydrophobic intramolecular interactions with helices α_7' and α_8' (red) from an adjacent monomer. In (b) helix α_8cF (teal) forms extensive intramolecular hydrophobic interactions with helices α_8dF (teal) and α_7 (green). The solvent-exposed polar residues Asp592, Glu593 and Lys596 of helix α_7 [positions -12, -11 and -8; bottom panel in (b)] in *T. thermophilus* MS are replaced by the hydrophobic residues Val525, Leu526 and Ser529 [positions -12, -11 and -8; bottom panel in (a)] that form part of the intramolecular interface in *T. maritima* MS. The solvent-exposed Pro611, Gln614 and Glu617 residues [positions 7, 10 and 13; bottom panel in (b)] in *T. thermophilus* MS are replaced by the hydrophobic residues Leu544, Thr547 and Leu552 [positions 7, 10 and 13; bottom panel in (a)] that reside in helix α_8 and form part of its intermolecular interface with α_7 and α_1 in *T. maritima* MS. In *T. thermophilus* MS, the hydrophobic residues Ala623 and Leu626 residing in α_8cF [positions 19 and 22; bottom panel in (b)] are replaced by the charged residues Glu556 and Glu559 in *T. maritima* MS.

(i) In the Fol modules with a classical fold, three outward-facing hydrophobic residues (in positions -12, -11 and -8; Fig. 7a) in helix α_7 which form intramolecular contacts with the α_7' and α_8' helices of an adjacent monomer are all substituted by charged residues in the Fol modules with a $\beta_8\alpha_7$ fold (Fig. 7b).

(ii) Similarly, in the $\beta_8\alpha_8$ modules three inward-facing hydrophobic residues in helix α_8 that form part of the interface with the adjacent α_7 and α_1 helices (in positions 7, 10 and 13; Fig. 7a) and two outward-facing residues (in positions 16 and 17; Fig. 7a) which form intramolecular contacts with the α_7' helix of an adjacent monomer are mainly replaced by charged residues in $\beta_8\alpha_7$ modules (Fig. 7b).

(iii) In $\beta_8\alpha_8$ modules, there is a cluster of mainly charged residues downstream of the α_8 helix (in positions 19, 22 and 23; Fig. 7a) that are all substituted with hydrophobic residues in $\beta_8\alpha_7$ modules. The latter residues that reside on the α_8cF helix are part of an extensive hydrophobic interface formed between helices α_8cF , α_8dF and α_7 (Fig. 7b).

(iv) In Fol modules with a $\beta_8\alpha_7$ fold helices α_6 and α_7 are longer than in modules with a $\beta_8\alpha_8$ fold (Fig. 2b), and these extensions form part of the extended interface with helices α_8cF and α_8dF (Supplementary Fig. S3).

A phylogenetic analysis (Supplementary Fig. S4) demonstrating that the *T. maritima* Fol module groups with other members of the MeTr family and not with the MS Fol modules may lend support to the notion that *T. maritima* MS is an outlier with a distinct method of connecting the Fol module to the Cob module. More structures of these modules are required to ascertain the significance of the observed variance. It is tempting to speculate that the variation in the TIM-barrel fold and the unique C-terminal arrangement in the majority of the Fol modules in MSs may have functional implications. Given the ability of the linker between the Fol and Cob modules to propagate structural changes, we hypothesize that this linker plays a significant role in the structural rearrangements that are required to accom-

modate the multitude of conformations observed during catalysis by and reactivation of MS (Fig. 1a).

Despite the clear structural differences in the C-terminus of the various MS Fol modules, the rest of the Fol module (including the folate-binding site) is essentially indistinguishable across all structures (Figs. 3 and 5; Evans *et al.*, 2004; Doukov *et al.*, 2007; Sjuts *et al.*, 2015). In general, the interactions between the folate substrate and active-site residues are mostly invariable in all folate-corrinoid methyltransferases. However, small differences can be found, as illustrated by the comparison of the CH₃-H₄folate-binding site for selected representative methyltransferases shown in Fig. 5.

While the structures of the excised Fol modules do not reveal any substantial CH₃-H₄folate-induced structural changes (Supplementary Fig. S2), it cannot be ruled out that such changes may occur upon the formation of a ternary complex. Structural changes could be induced upon the addition of the Cbl cofactor and its binding module in the context of the full-length enzyme. This ternary catalytically active conformation would support methyl transfer from the CH₃-H₄folate substrate to the cob(I)alamin cofactor. However, comparison of our *T. thermophilus* MS structures with those of other corrinoid-binding methyltransferases suggest that during catalysis the folate substrate is most likely to be activated by subtle changes in the active site. Moreover, very limited motion of the folate substrate is expected upon its interaction with the corrinoid cofactor during catalysis. While there is an emerging picture of the binary structures of protein-CH₃-H₄folate complexes, these might solely represent a binding (or recognition) mode and not an activation, catalysis-ready mode. The differences between the binding and the activation modes during the catalytic steps warrant future investigation.

5. Conclusions

We determined the crystal structure of the MS Fol module from *T. thermophilus* in the presence and absence of the CH₃-H₄folate substrate. We find that the active-site arrangement and the mode of folate recognition and binding are indistinguishable from those of previously determined folate-dependent methyltransferases, with the exception of MT2DH. Despite the similarities, our structures of the Fol module from *T. thermophilus* MS reveal a unique variation in the C-terminal region distinct from other folate-binding proteins with a classical $\beta_8\alpha_8$ fold. This unique topology is present in other MS Fol modules for which structures have been deposited. The fold of these MS Fol modules is a modified version of the classical TIM-barrel fold and can be best described as a $\beta_8\alpha_7$ barrel. Critical for the formation of this TIM barrel with a 'twist' is the presence on its C-terminus of residues that are part of a linker region connecting the Fol module to the Cob module in MS enzymes. We propose that the presence of this unique $\beta_8\alpha_7$ fold in the Fol module of MS proteins distinguishes them from other members of the MeTr family.

Acknowledgements

The authors acknowledge GM/CA at the Advanced Photon Source for beam time and thank Clay Brown (LSI, University of Michigan) for providing the pMCSG7 plasmid. KY carried out molecular biology, protein expression and crystallography; KY and MK analysed the data. Both authors contributed to the manuscript. The authors declare no competing financial interests.

Funding information

This work was supported by the American Heart Association (grant 13SDG14560056 to MK).

References

- Achari, A., Somers, D. O., Champness, J. N., Bryant, P. K., Rosemond, J. & Stammers, D. K. (1997). *Nature Struct. Biol.* **4**, 490–497.
- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst. D* **58**, 1948–1954.
- Banerjee, R. (1999). Editor. *Chemistry and Biochemistry of B₁₂*. New York: Wiley.
- Bommer, M., Kunze, C., Fessler, J., Schubert, T., Diekert, G. & Dobbek, H. (2014). *Science*, **346**, 455–458.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst. D* **66**, 12–21.
- Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. (2004). *Genome Res.* **14**, 1188–1190.
- Doukov, T., Hemmi, H., Drennan, C. & Ragsdale, S. (2007). *J. Biol. Chem.* **282**, 6609–6618.
- Doukov, T., Seravalli, J., Stezowski, J. J. & Ragsdale, S. W. (2000). *Structure*, **8**, 817–830.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486–501.
- Evans, J., Huddler, D., Hilgers, M., Romanchuk, G., Matthews, R. & Ludwig, M. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 3729–3736.
- Goldman, A. D., Beatty, J. T. & Landweber, L. F. (2016). *J. Mol. Evol.* **82**, 17–26.
- Hampele, I. C., D'Arcy, A., Dale, G. E., Kostrewa, D., Nielsen, J., Oefner, C., Page, M. G. P., Schönfeld, H.-J., Stüber, D. & Then, R. L. (1997). *J. Mol. Biol.* **268**, 21–30.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Huang, S., Romanchuk, G., Patridge, K., Lesley, S., Wilson, I., Matthews, R. & Ludwig, M. (2007). *Protein Sci.* **16**, 1588–1595.
- Kabsch, W. (2010). *Acta Cryst. D* **66**, 125–132.
- Koutmos, M., Pejchal, R., Bomer, T. M., Matthews, R. G., Smith, J. L. & Ludwig, M. L. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 3286–3291.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Kung, Y., Ando, N., Doukov, T. I., Blasiak, L. C., Bender, G., Seravalli, J., Ragsdale, S. W. & Drennan, C. L. (2012). *Nature (London)*, **484**, 265–269.
- Marchler-Bauer, A. *et al.* (2015). *Nucleic Acids Res.* **43**, D222–D226.
- Matthews, R., Koutmos, M. & Datta, S. (2008). *Curr. Opin. Struct. Biol.* **18**, 658–666.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Metz, J. (1992). *Annu. Rev. Nutr.* **12**, 59–79.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst. D* **67**, 355–367.
- Payne, K. A., Quezada, C. P., Fisher, K., Dunstan, M. S., Collins, F. A., Sjuts, H., Levy, C., Hay, S., Rigby, S. E. & Leys, D. (2014). *Nature (London)*, **517**, 513–516.
- Ragsdale, S. W. & Pierce, E. (2008). *Biochim. Biophys. Acta*, **1784**, 1873–1898.
- Schrauzer, G. N., Deutsch, E. & Windgassen, R. J. (1968). *J. Am. Chem. Soc.* **90**, 2441–2442.
- Sjuts, H., Dunstan, M. S., Fisher, K. & Leys, D. (2015). *Acta Cryst. D* **71**, 1900–1908.
- Smith, A. E. & Matthews, R. G. (2000). *Biochemistry*, **39**, 13880–13890.
- Stols, L., Gu, M., Dieckman, L., Raffin, R., Collart, F. R. & Donnelly, M. I. (2002). *Protein Expr. Purif.* **25**, 8–15.
- Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
- Watkins, D. *et al.* (2002). *Am. J. Hum. Genet.* **71**, 143–153.
- Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R. & Kuramitsu, S. (2000). *Nature Struct. Biol.* **7**, 943–945.