Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a closed-form uroporphyrinogen-III *C*-methyltransferase from *Thermus thermophilus*

Peter H. Rehse, Tomoe Kitao and Tahir H. Tahirov*

APCG, RIKEN Harima Institute, 1-1-1 Kouto, Mikazuki-cho, Sayo-gun, Hyogo 679-5148, Japan

Correspondence e-mail: tahir@spring8.or.jp

Received 8 February 2005
Accepted 18 March 2005

PDB References:
uroporphyrinogen-III C-methyltransferase, 1va0, r1va0sf; cofactor complex, 1v9a, r1v9asf.

Uroporphyrinogen-III *C*-methyltransferase from *Thermus thermophilus* is a multifunctional protein responsible for two of the eight *S*-adenosyl-methionine-dependent methylations of the corrin ring during vitamin B_{12} synthesis. The structure of this protein has been solved to 2.0 Å resolution in both the apo and cofactor-bound form. The monomer consists of two domains, *A* and *B*, each consisting of a five-stranded β-sheet and two or three α-helices, with the cofactor bound at the interface. The biological unit is the dimer found in the asymmetric unit. This dimer is related by a non-crystallographic twofold such that two *B* domains combine to form a long ten-stranded β-sheet. When compared with solved related structures, this structure shows clear differences in the region involved in cofactor and substrate binding, affirming the role of several previously implicated residues and questioning others. The solved related structures are characterized by an exposed active site. The *T. thermophilus* structure has this site restricted by the interaction of a flexible loop structure with a highly conserved residue, suggesting a mechanistic role. This structure represents the 'closed' form of the protein.

1. Introduction

The biosynthesis of cobalamin (vitamin B_{12}) is an incredibly complex process involving a large number of enzymatic steps (Debussche *et al.*, 1993). Central to this is the *S*-adenosyl methionine (SAM) dependent addition of eight methyl groups to the corrin ring, starting from uroporphyrinogen-III. The enzymes are related through sequence and structure, with some able to catalyze multiple steps. In bacterial species, the biosynthetic pathways responsible for these methylations diverge into independent aerobic and anaerobic pathways before converging again to give the final product (Raux *et al.*, 1996). The anaerobic pathway is dependant on cobalt binding. The point of divergence of the biosynthetic pathway into aerobic and anaerobic differs between different bacterial species, with the earliest occurring after the production of precorrin-2 (Schubert *et al.*, 1998).

Uroporphyrinogen-III *C*-methyltransferase from *Thermus thermophilus* HB8 (ttSUMT) catalyzes the first two SAM-dependent methylations (uroporphyrinogen-III $\rightarrow$ precorrin-1 $\rightarrow$ precorrin-2; Fig. 1) and hence can be considered to be a multifunctional protein. However, the symmetric nature of the uroporphyrinogen-III structure suggests that the active-site requirements of the two reactions are essentially the same.

Recently, the X-ray structure and mutagenesis of uroporphyrinogen-III *C*-methyltransferase from *Pseudomonas denitrificans* (pdSUMT) was solved to 2.7 Å resolution

(Vévodová *et al.*, 2004). This protein shares 38% identity with ttSUMT.

The structure of a much larger protein (CysG) performing a related function in *Salmonella enterica* was solved by Stroupe *et al.* (2003). However, this enzyme is truly multifunctional in that it additionally acts as an NAD^+-dependent dehydrogenase, converting precorrin-2 to sirohydrochlorin, followed by a ferrochelatase function. The CysG protein has been shown to catalyze the insertion of cobalt (Fazzio & Roth, 1996) as does the homologous protein from *Escherichia coli* (Spencer *et al.*, 1993), even though the latter organism does not produce cobalamin. The latter two functions are part of a separate enzymatic module from the SAM-dependent methyltransferase function and were shown to function independently (Warren *et al.*, 1994). The methyltransferase portion of this protein has 39% identity to ttSUMT.

A functionally related enzyme from *Bacillus megaterium* has also been solved (Schubert *et al.*, 1998). This cobalt-precorrin-4 methyltransferase (CbiF) is responsible for the fifth SAM-dependent methylation of the corrin ring, converting cobalt-precorrin-4 to cobalt-precorrin-5. Apart from the additional methyl groups, the substrate differs from uroporphyrinogen-III in that it is complexed with Co^{2+}. It has 28% identity to ttSUMT. The sequence length of 234 amino acids is similar to that of ttSUMT.

Here, we present the X-ray crystallographic structure of uroporphyrinogen-III *C*-methyltransferase from *T. thermophilus* HB8 solved to 2.0 Å resolution in both the apo form and in complex with its demethylated cofactor *S*-adenosyl homocysteine (SAH). We discuss this structure in relationship to other proteins involved in corrin-ring methylations during the synthesis of vitamin B_{12}.

2. Materials and methods

2.1. Protein expression

The polymerase chain reaction (PCR) was used for gene amplification of *T. thermophilus* HB8 genomic DNA. The PCR product was ligated with pT7blue (Novagen) and digested with *Nde*I and *Bgl*II. The fragment was inserted into the expression vector pET-11a made linear by digestion with *Nde*I and *Bam*HI and transformed into *E. coli* strain BL21 (DE3). The cells were grown for 20 h at 310 K in 9.0 l medium containing 50 μg ml^{-1} ampicillin. Cells (22.6 g) were harvested by centrifugation at 6500 rev min^{-1} for 5 min and suspended in 46 ml 20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 5 m*M* 2-mercaptoethanol and 1 m*M* phenylmethylsulfonyl fluoride. The cells were disrupted by sonication followed by heat treatment at 343 K for 13 min. *T. thermophilus* proteins are not expected to denature at this temperature. The cell debris and denatured proteins were removed by centrifugation (15 000 rev min^{-1}, 30 min, 277 K).

2.2. Purification

The supernatant was applied onto a HiPrep 26/10 desalting column (53 ml; Amersham Biosciences) using 20 m*M* Tris–HCl pH 8.0 (buffer *A*). The elutant was applied onto a SuperQ Toyopearl 650M (30 ml; Tosoh) column equilibrated in buffer *A* and eluted with a 0–0.3 *M* NaCl linear gradient. The main protein peak was desalted using a HiPrep 26/10 column (Amersham Biosciences) with buffer *A* and applied onto a Resource Q (6 ml; Amersham Biosciences) column equilibrated in buffer *A* and eluted with a 0–0.3 *M* NaCl gradient. The main protein fraction was desalted with a HiPrep 26/10 column equilibrated in 10 m*M* sodium phosphate pH 7.0 (buffer *B*), applied onto a CHT20-I (20 ml; Bio-Rad) column and eluted with a 0.01–0.2 *M* sodium phosphate gradient in buffer *B*. The main protein peak was then concentrated and applied onto a HiLoad 16/60 Superdex 200 (120 ml; Amersham-Biosciences) column equilibrated with 20 m*M* Tris–HCl, 200 m*M* NaCl pH 8.0. The purified protein (7.6 mg) was homogenous on SDS–PAGE. The protein was concentrated to 33 mg ml^{-1} using ultrafiltration (Amicon; 5 kDa cutoff). The extinction coefficient was determined to be 18 900 cm^{-1} M^{-1}; the calculated molecular weight is 26 233 Da.

2.3. Dynamic light-scattering measurements

Dynamic light-scattering measurements of the purified protein were performed using DynaPro MS/X (Protein Solution). Protein samples (1.0 mg ml^{-1}) were prepared in a solution of 20 m*M* Tris–HCl buffer pH 8.0 containing 200 m*M* NaCl. Protein samples were centrifuged at 10 000 rev min^{-1} for 10 min before measurement. The measurements were performed at 291 K. All data were analyzed using *DYNAMICS* software (Protein Solutions). The protein solution was monodisperse and yielded a molecular weight of

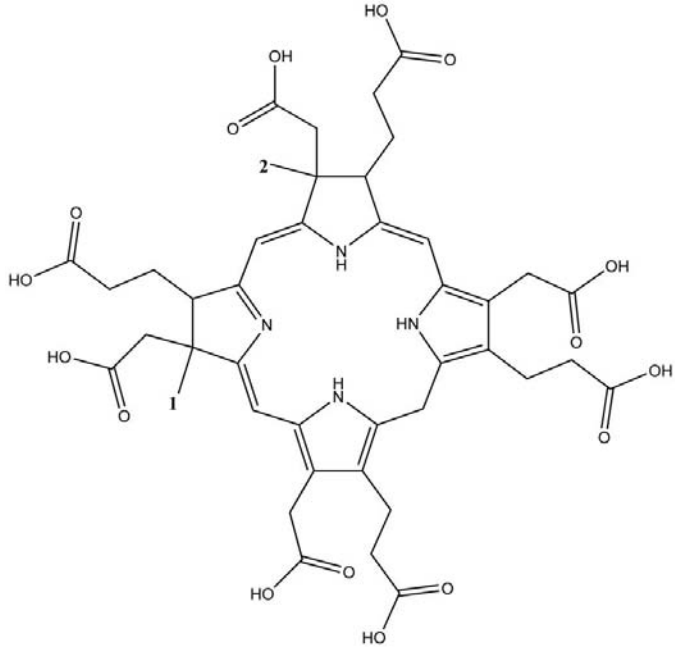

Figure 1
Structure of precorrin-2, the product of SAM-dependent uroporphyrinogen-III *C*-methyltransferase. The two independent methylation sites are designated 1 and 2. Methylation only at site 1 generates the intermediate precorrin-1.

53 kDa. Using the calculated molecular weight, this suggests that the protein is a dimer in solution.

2.4. Crystallization

Crystals were initially obtained by the microbatch method (Chayen *et al.*, 1990) using a TERA crystallization robot and a screening kit designed for high-throughput protein crystallization (Sugahara & Miyano, 2002). Using 100 m*M* citrate pH 5.3 and 7.5% polyethylene glycol (PEG) 20K, crystals appeared as blocks of 0.2 × 0.2 × 0.2 mm in size after about two months, but were not reproducible under the above conditions or using the sitting-drop method mixing 0.5 µl protein solution with 0.5 µl reservoir solution. In the latter case only needle crystals were obtained. The crystallization temperature was 295 K. Using freshly prepared sample, similar crystals appeared after 2 d using the hanging-drop method. Further conditions were explored by varying buffer type and pH, various salts and additives and precipitant type and concentration in order to improve the resolution and problems of fragility; however, the best crystals were obtained varying only the pH and PEG concentration. Three distinct crystal forms were recognizable, often growing in the same drop, with the tetragonal bipyramidal and tetragonal pyramidal forms being extremely fragile especially when exposing to various cryoprotectants. The third plate-like form was far less fragile and withstood short exposure to cryoprotectants. Although the mosaicity was reasonable, the diffraction was limited to around 2.5 Å. Fragments of the two other forms could be mounted; however, with the tetragonal bipyramidal crystals the diffraction pattern was always fragmented or had high mosaicity. The maximum resolution obtained was around 2.6 Å. The data sets described below were from a mounted fragment of the tetragonal pyramidal crystals. The fragments were quite large and easily separated.

2.5. SAM soaking

A 1 m*M* solution of SAM was prepared by mixing 1 µl 10 m*M* SAM stock (in water) with 9 µl reservoir solution. The protein solution surrounding the crystals was removed and replaced by 1 m*M* SAM solution. The crystals were left for 18 h at 295 K before mounting.

2.6. Data collection

The buffer surrounding the crystals was changed to a cryoprotectant buffer consisting of 100 m*M* citrate pH 5.3, 12% PEG 20K and 30% ethylene glycol through the initial addition of reservoir solution (8 µl) followed by increasing amounts of cryoprotectant buffer (2–5 µl) coupled with removal of excess drop solution. The crystal fragments were separated, mounted on cryoloops and then flash-frozen under an N_2 stream at 100 K. It was important to effect this buffer exchange over less than 10 min.

A data set for the apo form was collected from a single crystal on an R-AXIS VII diffractometer using Cu $K\alpha$ radiation. The data were collected as 185 0.5° oscillation images. The crystal-to-image plate distance was 145 mm, with a 15 min exposure. The crystal mosaicity was determined to be 0.28°.

Table 1
Processing and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Apo	Cofactor-bound
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 60.50, b = 63.57, c = 131.90	a = 59.98, b = 63.81, c = 132.42
Processing		
Resolution (Å)	31.78–1.97 (2.04–1.97)	41.5–1.93 (2.00–1.93)
R_{merge} (%)	3.9 (49.9)	7.2 (48.6)
$I/\sigma(I)$	18.8 (1.65)	15.1 (3.03)
Completeness (%)	100.0 (99.9)	99.6 (100.0)
Multiplicity	3.60 (3.53)	3.97 (4.10)
Refinement		
Resolution (Å)	31.78–1.97 (2.09–1.97)	41.5–2.00 (2.07–2.00)
Reflections	35288	33592
Reflections in R_{free}	1751 (261)	1678 (162)
Protein atoms	3483	3450
Solvent atoms	244	209
Other atoms	23	78
R factor (%)	19.9 (26.9)	21.5 (28.7)
R_{free} (%)	22.7 (30.3)	24.6 (30.7)
Model quality		
Residues in		
Most favored region	93.6	92.4
Additionally allowed region	6.2	7.6
Generously allowed region	0.3	0.0
Other regions	0.0	0.0
R.m.s. deviations		
Bonds (Å)	0.009	0.009
Angles (°)	1.60	1.60
Average B factors (Å^2)		
Main chain	38.1	37.3
Side chain	44.7	43.5
Solvent	49.1	46.7
Others	57.0	68.6
Ligand	—	27.5
Overall	41.9	40.7

A data set from the cofactor-bound form was collected and processed in a similar way at the synchrotron beamline BL-26a at SPring-8 (Harima, Japan). The data were collected as 100 1.0° oscillation images. Exposure times were 55 s and the crystal-to-film distance was 200 mm. The crystal mosaicity was determined to be 0.25°. Data were indexed, integrated and scaled with *DENZO* and *SCALEPACK* implemented in the *HKL*2000 program package (Otwinowski, 1993; Otwinowski & Minor, 1997). The resolution cutoff was chosen as the edge of the imaging plate where the data was complete and the R_{merge} of the highest resolution shell remained below 0.5. Data collection is summarized in Table 1; based on systematic absences the space group was determined to be $P2_12_12_1$.

2.7. Structure determination and model refinement

The structure of the apo form was solved by molecular replacement using *MOLREP* (Collaborative Computational Project, Number 4, 1994; Vagin & Teplyakov, 1997) using the biological unit dimer of cobalt-precorrin-4 methyltransferase from *B. megaterium* (PDB code 2cbf; Schubert *et al.*, 1998). The cofactor-bound form was solved using the partially refined

apo form as the molecular-replacement model. Refinement, including simulated annealing and non-crystallographic symmetry in the initial stages, was carried out using *CNS* (Brünger *et al.*, 1998; see Table 1). The structure was built and modified using the program *QUANTA* (Accelrys). Stereochemical analysis of the structure was performed using *PROCHECK* (Laskowski *et al.*, 1993). Figures were generated with *MOLSCRIPT* (Kraulis, 1991), *RASTER3D* (Merritt & Bacon, 1997) and *GRASP* (Nicholls *et al.*, 1991). Sequence alignments were performed using *CLUSTALW* (Thompson *et al.*, 1994).

3. Results and discussion

3.1. Overall structure of apo form

The structure was solved using the molecular-replacement method and CbiF (Schubert *et al.*, 1998) as the probe molecule. Uroporphyrinogen-III *C*-methyltransferase (SUMT) converts uroporphyrinogen-III to precorrin-2 by a two-step methylation involving two SAM cofactors. Two further methylations produce precorrin-4, which when bound to cobalt becomes cobalt-precorrin-4. It is this substrate which the probe molecule methylates, also using SAM as cofactor. Although the homology was only 28%, it was felt that the reaction mechanism was similar enough that molecular replacement could be attempted. At the time, the structures of pbSUMT and CysG were unavailable. The crystal space group was $P2_12_12_1$, with unit-cell parameters a = 60.50, b = 63.57, c = 131.90 Å and two 240-residue molecules expected in the asymmetric unit. Molecular replacement using the CbiF monomer failed to yield a clear solution. However, when the postulated biological CbiF dimer was used a solution was obtained. The final structure contained two molecules in the asymmetric unit, of which *A*57–60, *B*53–62, *A*227–240 and *B*239–240 could not be built into density. The model contained five ethylene glycols, three chloride ions and 244 water molecules. The terminal two atoms of the Lys*A*93 side chain were missing, as was the side chain beyond C^{β} for Leu*B*36. An alternate conformation was built for Glu*A*27. Refinement statistics are given in Table 1. Ser*B*190 was the sole residue in the generously allowed region defined by *PROCHECK* (Laskowski *et al.*, 1993). Pro214 was a *cis*-proline.

The core of both domains consists of a five-stranded β-sheet fronted on one side by two α-helices and on the other by a less distinct arrangement of helices. The β-sheet of domain *A* is completely parallel and composed of non-consecutive β-strands ($\beta3\beta2\beta4\beta1\beta5$), while the primary sequence arrangement is ($\beta\alpha\beta\alpha\beta\alpha\beta\alpha\beta$). In domain *B* the β-sheet is primarily antiparallel and both the ordering of β-strands ($\beta6\beta7\beta10\beta8\beta9$) and the primary sequence arrangement ($\beta\beta\alpha\beta\beta\alpha\beta$) differ. In domain *B*, $\beta6$ and $\beta7$ are parallel. The prominent α-helices to one side of each domain correspond to the last α-helices of each domain: $\alpha3$ and $\alpha4$ in domain *A* and $\alpha5$ and $\alpha6$ in domain *B*. The two domains are joined by a long essentially random coil which contains a short 3_{10}-helix ($\eta2$) near the middle. Two further 3_{10}-helices are found: one in the loop joining $\beta1$ and $\alpha1$ ($\eta1$) and a more pronounced one immediately following the final β-sheet ($\eta3$). The interactions between the two domains are not particularly strong, involving loops $\beta8\beta9$ and $\beta1\eta1$ in one instance and $\beta3\alpha3$ and $\beta6\beta7$ in another. The loop $\beta4\alpha4$ and $\eta2$ interact to a small degree with the β-sheet of domain *B*, as does the turn around residue 141 with the loop $\beta3\alpha3$. Fig. 2 shows the corresponding secondary-structure elements with the primary sequence. Fig. 3(*a*) shows the monomer structure of the protein orientated such that the similarity between the two domains and their lack of interaction is obvious.

The biological unit is produced by a non-crystallographic twofold such that the N-terminal half of $\beta6$ forms an antiparallel β-sheet with itself. Fig. 3(*b*) shows the dimer looking down the resultant ten-stranded β-sheet or perpendicular to the twofold axis. The orientation of the monomer domains are stabilized by the dimer interactions. Along with dynamic light-scattering measurements, these observations support the dimer as the biological unit.

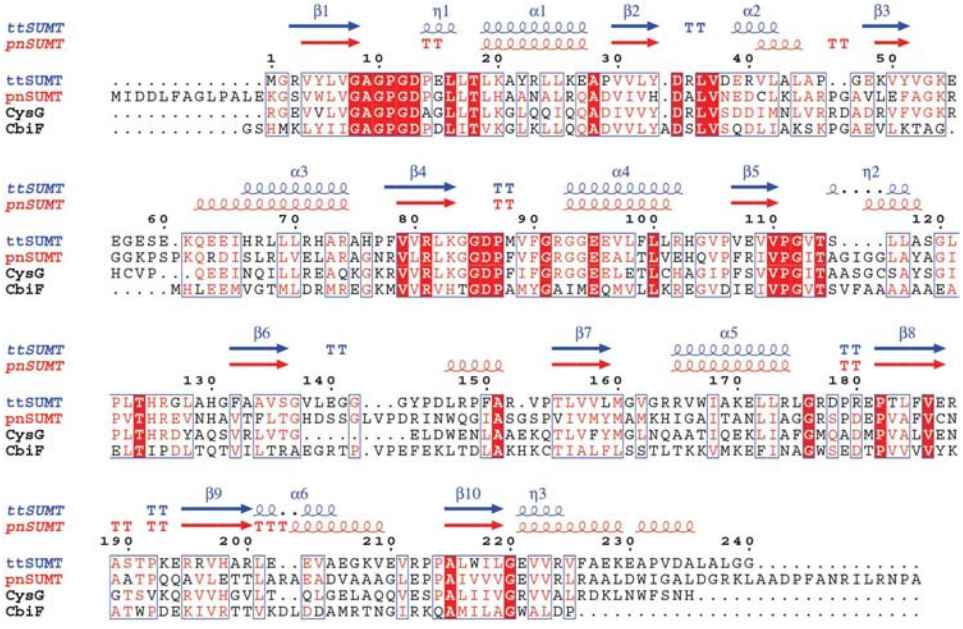

Figure 2
Sequence alignment of uroporphyrinogen-III *C*-methyltransferase from *T. thermophilus* (ttSUMT) and from *P. denitrificans* (pdSUMT) with cobalt-precorrin-4 methyltransferase from *B. megaterium* (CbiF), used as the molecular-replacement probe for the *T. thermophilus* structure, and the CysG protein (residues 215–457) from *S. enterica*. Secondary-structure elements from the respective proteins were obtained using *DSSP* (Kabsch & Sander, 1983). The figure was produced using *ESPript* (Gouet *et al.*, 1999). Alignment was performed using *CLUSTALW* (Thompson *et al.*, 1994).

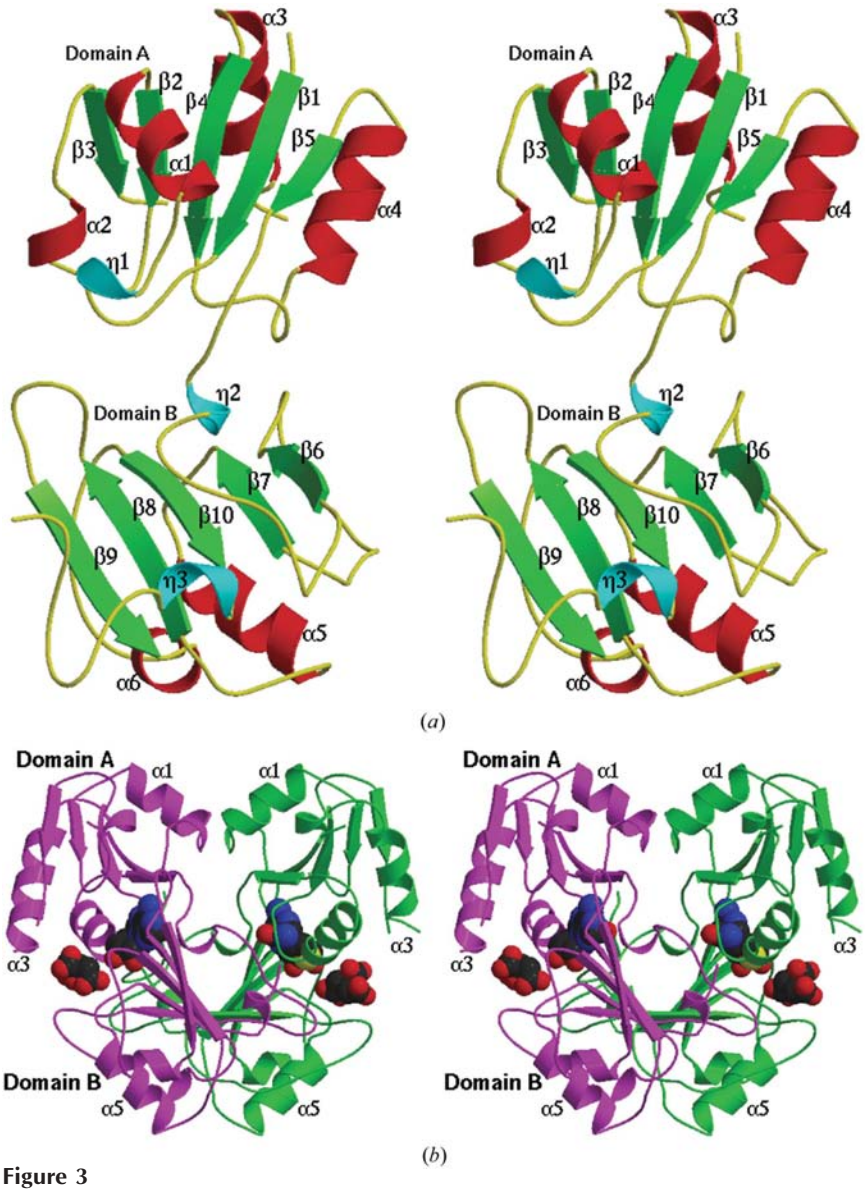

Figure 3
Stereo secondary-structure cartoons of uroporphyrinogen-III *C*-methyltransferase from *T. thermophilus*. (*a*) Monomer orientated to show the similarity between the two domains and their lack of interaction. The α-helices are red, β-sheets green and 3_{10}-helices blue. (*b*) Biological unit in complex with cofactor and citrate molecule. Labels refer to the secondary-structure elements described in Fig. 2.

3.2. Cofactor-bound structure

The overall structure of the cofactor-bound form was essentially the same as the apo form. The final structure was missing residues *A*1, *B*1, *A*55–60, *B*55–63, *A*227–240 and *B*239–240. The structure contained one SAH and one citrate molecule per monomer and 209 water molecules. The entire side chain of Lys54 was missing for both the *A* and *B* chain, as was the side chain beyond C^{β} for Glu*A*63. The three terminal side-chain atoms of Arg*A*224 and Glu*B*39 are also missing. Refinement statistics are given in Table 1. As in the apo structure, Pro214 was a *cis*-proline. Fig. 3(*b*) shows the placement of the cofactor within the overall fold and also the placement of the citrate molecules. Based on the orientation of the homocysteine, the citrate molecule, although smaller, occupies the site near where the substrate is expected to bind. The residues surrounding the SAH and citrate molecules are given in Fig. 4.

A comparison of the cofactor binding of ttSUMT with that of CysG (Stroupe *et al.*, 2003), pdSUMT (Vévodová *et al.*, 2004) and CbiF (Schubert *et al.*, 1998) shows the orientation of the SAH molecule to be nearly identical, although there are some marked differences with respect to the surrounding residues. It appears that the SAM moiety in the ttSUMT structure has undergone a spontaneous or enzyme-catalyzed conversion to its demethylated form (SAH). In the previously solved structures SAH is the bound cofactor, not SAM.

In the CysG protein the Tyr381 side chain is in hydrophobic contact with the cofactor SD and is stacked against the aromatic ring of Phe307. These residues, along with Met382, form a hydrophobic pocket in the region where methyl transfer is to occur. This relationship is maintained in the pdSUMT structure (Phe183, Met184, Tyr106). In the CbiF structure the methionine is replaced by Leu183, but the aromatic ring stacking is maintained (Phe183, Leu184, Tyr107). In the ttSUMT structure an analogous residue to Tyr381 (CysG) does not exist, the closest residue being Leu159. This particular residue is very interesting since the mutation Y183A in pdSUMT abolishes activity (Vévodová *et al.*, 2004). Its absence allows the Met160 (ttSUMT) residue to move such that its amide is able to form a hydrogen bond with both O2 and O3 of the SAH ribose moiety. In the pdSUMT and CysG structures it can only form a hydrogen bond with O2. In both structures the methionine SD is in hydrophobic contact with the SAH SD. Extra hydrogen bonds to the SAH molecule in the ttSUMT structure are provided by the amide of Ser115 and the carbonyl O atom of Ala189. A similar relationship exists in the other solved structures. In the CysG structure there are two further hydrophobic contacts to the SAH molecule provided by Cys336 and Leu438. These two residues are also in contact with each other. There is no analogous residue to Cys336 in the other solved structures, although in the CbiF structure Ala136 is nearby. Leu438 (CysG) corresponds to Ile242 in the pdSUMT structure and Met242 in CbiF. In the ttSUMT structure the corresponding residue is Leu216, but it is shifted closer towards where the cysteine would be. Generally speaking, the region surrounding the cofactor binding is less hydrophobic in the ttSUMT structure than the CycG structure, with the latter more similar to the pdSUMT structure.

When the CbiF protein is compared, further differences become obvious. The replacement of Met160 (ttSUMT) by Leu184 maintains the hydrogen bonds to O2 and O3 of the ribose ring seen in the ttSUMT structure, but the hydrophobic contact of the side chain to the SAH sulfur is lost. In pdSUMT, mutation of this methionine to alanine abolishes the activity (Vévodová *et al.*, 2004). The demethylation of the SAM cofactor occurs in this region. Generally speaking, however, the degree of hydrogen bonding and hydrophobic contacts is most similar between CbiF and ttSUMT.

3.3. Fold comparison and substrate access

The aligned backbones of the four structures are shown in Fig. 5(*a*), with the absolutely conserved residues displayed in full. Both the residues and cofactor are from the ttSUMT structure described here. Although the majority of the fold appears to be well maintained, it is clear that the conserved residues do not cluster near the area where substrate is expected to come into contact with the SAM cofactor. Two loops (loops *A* and *B* in Fig. 5*a*) appear to be important in the area of substrate binding. Although neither of the loops contain absolutely conserved residues, loop *A* has a highly conserved main-chain structure. Loop *B* (residues 137–148) is much more flexible and its role may be instrumental in regulating substrate access. In the structures of CysG (Stroupe *et al.*, 2003), pdSUMT (Vévodová *et al.*, 2004) and CbiF (Schubert *et al.*, 1998) this loop does not prevent access to the methylation site of the SAM cofactor, whereas in the ttSUMT structure it penetrates much further into the cleft, to the point of interacting with the second domain (see Fig. 5*a*). The corresponding loop from pdSUMT is three residues longer than that found in the ttSUMT structure. Fig. 5(*b*) shows the structure of pdSUMT orientated to show the large open cleft. The ttSUMT structure is shown in Fig. 5(*c*) orientated to demonstrate the domain-bridging loop. In ttSUMT, the side chain of Glu140 interacts with the side chain of Arg92 in the second domain. The latter residue is invariant in the uroporphyrinogen-III-binding proteins, but is an alanine in CbiF. The Arg92 side chain in turn interacts with the Glu95 side chain, which is invariant in all the sequences shown in Fig. 2. The invariant nature of this latter

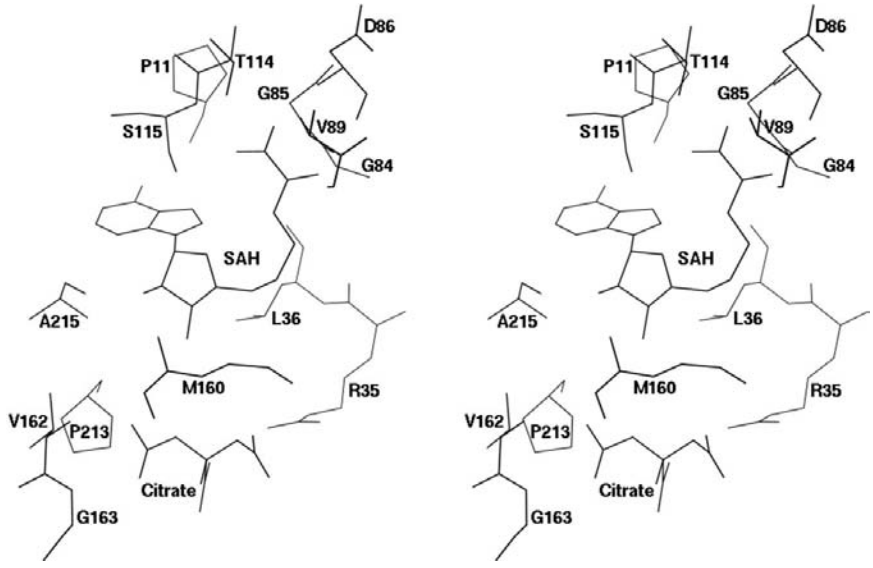

Figure 4
Stereo diagram of the cofactor-binding site of uroporphyrinogen-III *C*-methyltransferase from *T. thermophilus* (ttSUMT).

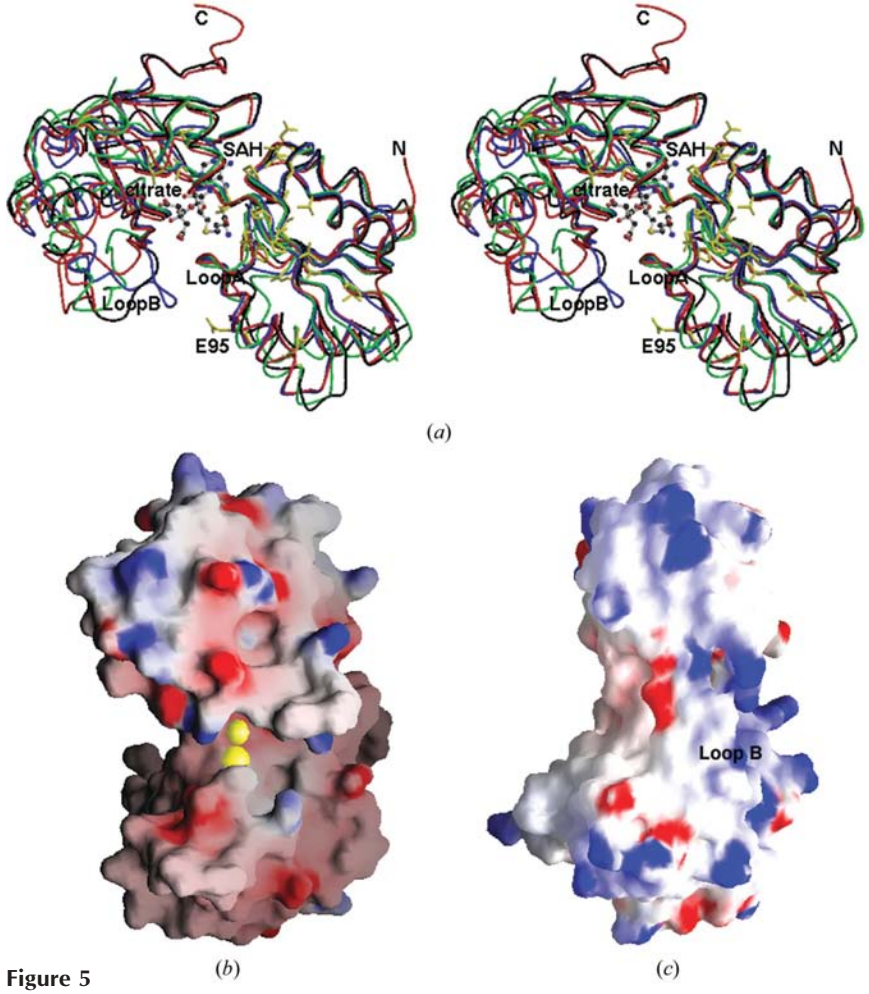

Figure 5
Effect of loop *B*: the 'closed' structure. (*a*) Stereoview of overlapped backbones of ttSUMT (blue), pdSUMT (red), CbiF (green) and residues 215–457 of the CysG protein (black). Residues absolutely conserved between all four are displayed in yellow. The cofactor, citrate and displayed residues are from the ttSUMT structure. (*b*) pdSUMT structure orientated to show the accessible cleft. (*c*) ttSUMT structure showing the domain-bridging loop *B*. Yellow spheres are the SD atoms of the methionine (lower) and cofactor (upper).

residue suggests the above interactions have biological significance. Although the loop does not block off the active-site cavity completely to solvent, the substrate is too large to be accommodated by the remaining space. Hence, the ttSUMT structure represents a 'closed' form of the enzyme and may reflect a means of regulating the activity of this enzyme.

3.4. Further active-site residues

Stroupe *et al.* (2003) discuss a number of residues important to substrate binding by CysG. They indicate that Arg298, Lys300 and Arg309 may be responsible for the binding of the acetic and propionic side chains of the reactive pyrrole ring. Arg298 corresponds to Arg81 in the ttSUMT structure and is conserved in all four structures. Lys300 and Arg309 of the CysG structure correspond to Lys83 and Arg92 in ttSUMT, respectively, and are conserved in the uroporphyrinogen-III-binding proteins but not in the CbiF structure. A number of nearby glutamic acid/glutamine residues are conserved. Glu95 and Glu96 (ttSUMT) are conserved throughout, but Gln62 (ttSUMT) is conserved in the uroporphyrinogen-III-binding proteins but not in the CbiF structure.

Stroupe *et al.* (2003) also suggest that Asn385 is important for substrate orientation of CysG, yet there is quite a lot of variation between the four structures. The citrate of the ttSUMT structure binds nearby. In the ttSUMT structure Arg35 forms a hydrogen bond to the citrate, yet there is no comparable residue in the two other structures. Stroupe *et al.* (2003) also indicate that Lys270 (CysG) is essential for catalysis, but in both the ttSUMT and CbiF structures there is no comparable residue. The positive charge provided by the loss of the lysine could be compensated for by Arg35 of the ttSUMT structure. Asp34 (ttSUMT) is maintained in all three structures with the side chain orientated in the same direction, yet the main-chain position varies with respect to the cofactor. Mutation of this residue restricted the ability of the pdSUMT to produce precorrin-2 even though precorrin-1 was produced (Vévodová *et al.*, 2004).

In the ttSUMT structure, Lys61 restricts access of the cofactor SD somewhat. In the pdSUMT structure, the analogous Lys77 is less restrictive. Similar restrictions are not found in the other structures described. The SE of Met160 (ttSUMT) also restricts access to the cofactor SD in the cofactor-bound version of the ttSUMT structure. In all other structures, including the ttSUMT apo form, the Met SE is rotated away and does not interfere.

4. Conclusion

A comparison of the three crystallographically solved structures allows one to make direct comparisons with respect to which residues are likely to be involved in the methylation reaction and which are more important to substrate binding. There is a large amount of structural variation in some parts of the binding cleft, yet in others there is strong conservation even though the substrate is different. Several attempts were made to bind substrate to the protein in the crystal, but these have proved to be unsuccessful. Of the three related enzymes whose structures have been previously reported, a highly flexible loop near the active site does not hinder the approach of the substrate to the methylation site. The reported ttSUMT structure has substrate access to this site prevented by the loop. This loop interacts with a highly conserved residue, suggesting biological significance of the 'closed state'.

We would like to acknowledge Seiki Kuramitsu and Shigeyuki Yokoyama for the plasmid vector used in the protein expression, and Yuki Sakaguchi for the technical help in the purification. This work was supported by a 'Protein 3000' Japan grant (Project TT0228/HTPF00122).

References

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Debussche, L., Thibaut, D., Cameron, B., Crouzet, J. & Blanche, F. (1993). *J. Bacteriol.* **175**, 7430–7440.
Fazzio, T. G. & Roth, J. R. (1996). *J. Bacteriol.* **178**, 6952–6959.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kraulis, J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Nicholls, A., Sharp, K. & Honig, B. (1991). *Proteins Struct. Function. Genet.* **11**, 281–296.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Raux, E., Lanois, A., Levillayer, F., Warren, M. J., Brody, E., Rambach, A. & Thermes, C. (1996). *J. Bacteriol.* **178**, 753–767.
Schubert, H. L., Wilson, K. S., Raux, E., Woodcock, S. C. & Warren, M. J. (1998). *Nature Struct. Biol.* **5**, 585–592.
Spencer, J. B., Stolowich, N. J., Roessner, C. A. & Scott, A. I. (1993). *FEBS Lett.* **335**, 57–60.
Stroupe, M. E., Leech, H. K., Daniels, D. S., Warren, M. J. & Getzoff, E. D. (2003). *Nature Struct. Biol.* **10**, 1064–1073.
Sugahara, M. & Miyano, M. (2002). *Tanpakushitsu Kakusan Koso*, **47**, 1026–1032.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vévodová, J., Graham, R. M., Raux, E., Schubert, H. L., Roper, D. I., Brindley, A. A., Scott, A. I., Roessner, C. A, Stamford, N. P. J., Stroupe, M. E., Getzoff, E. D., Warren, M. J. & Wilson, K. S. (2004). *J. Mol. Biol.* **344**, 419–433.
Warren, M. J., Bolt, E. L., Roessner, C. A., Scott, A. I., Spencer, J. B. & Woodcock, S. C. (1994). *Biochem. J.* **15**, 837–844.