

Elucidation of Substrate Specificity in the Cobalamin (Vitamin B₁₂) Biosynthetic Methyltransferases

STRUCTURE AND FUNCTION OF THE C20 METHYLTRANSFERASE (CbiL) FROM METHANOTHERMOBACTER THERMAUTOTROPHICUS^{*§}

Received for publication, May 9, 2007, and in revised form, June 13, 2007. Published, JBC Papers in Press, June 13, 2007, DOI 10.1074/jbc.M703827200

Stefanie Frank[‡], Evelyne Deery[‡], Amanda A. Brindley[‡], Helen K. Leech[‡], Andrew Lawrence[‡], Peter Heathcote[§], Heidi L. Schubert[¶], Keith Brocklehurst[§], Steve E. J. Rigby[§], Martin J. Warren^{‡,1}, and Richard W. Pickersgill^{§,2}

From the [‡]Protein Science Group, Department of Biosciences, University of Kent, Canterbury, Kent, CT2 7NJ, United Kingdom, the [§]School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, United Kingdom, and the [¶]Department of Biochemistry, University of Utah, Salt Lake City, Utah 84112-5650

Ring contraction during cobalamin (vitamin B₁₂) biosynthesis requires a seemingly futile methylation of the C20 position of the tetrapyrrole framework. Along the anaerobic route, this reaction is catalyzed by CbiL, which transfers a methyl group from *S*-adenosyl-L-methionine to cobalt factor II to generate cobalt factor III. CbiL belongs to the class III methyltransferases and displays similarity to other cobalamin biosynthetic methyltransferases that are responsible for the regiospecific methylation of a number of positions on the tetrapyrrole molecular canvas. In an attempt to understand how CbiL selectively methylates the C20 position, a detailed structure function analysis of the enzyme has been undertaken. In this paper, we demonstrate that the enzyme methylates the C20 position, that its preferred substrate is cobalt factor II, and that the metal ion does not undergo any oxidation change during the course of the reaction. The enzyme was crystallized, and its structure was determined by x-ray crystallography, revealing that the 26-kDa protein has a similar overall topology to other class III enzymes. This helped in the identification of some key amino acid residues (Asp¹⁰⁴, Lys¹⁷⁶, and Tyr²²⁰). Analysis of mutant variants of these groups has allowed us to suggest potential roles that these side chains may play in substrate binding and catalysis. EPR analysis of binary and ternary complexes indicate that the protein donates a fifth ligand to the cobalt ion via a gated mechanism to prevent transfer of the methyl group to water. The chemical logic underpinning the methylation is discussed.

Cobalamin (vitamin B₁₂) (Fig. 1*a*) is a modified tetrapyrrole that differs from other members of this fraternity, such as heme

and chlorophyll, in that the macrocycle has been contracted by loss of one of the bridging carbons of the molecular framework. Cobalamin biosynthesis is mediated via an elaborate pathway that requires the deployment of somewhere around 30 enzymes for its complete *de novo* synthesis (1–4). The situation is complicated further by the presence in nature of two similar but biochemically and genetically distinct syntheses (5, 6). These pathways are referred to as the aerobic or cobalt-late and anaerobic or cobalt-early routes (7). As their names suggest, the pathways differ in their requirement for molecular oxygen and the timing of cobalt insertion. The variance in the pathway is associated with the biosynthesis of the corrin ring component of cobalamin and in particular the events that see the transformation of the first macrocyclic intermediate uroporphyrinogen III into cobyrinic acid. This transformation includes the addition of eight *S*-adenosyl-L-methionine (SAM)³-derived methyl groups to the tetrapyrrole scaffold, although only seven are observed in the final product, since one is lost during the ring contraction process (Fig. 1*a*).

An interesting question to address is the rationale behind why there are so many methylations during cobalamin biosynthesis. The methylations are largely associated with prototrophic rearrangements and arguably evolved to prevent oxidation of the corrin as the earth's environment became more oxidizing (8). In so doing, some of the methylation processes, such as those associated with C17 and C1, have become associated with ring contraction and elimination of the extruded carbon fragment. An underappreciated yet pivotal reaction of the ring contraction process is methylation at C20. Although in principle it appears futile to methylate a carbon that is subsequently lost, chemical studies have shown that ring contraction only occurs if this position is derivatized into a tertiary alcohol (8) (Fig. 1*b*), and this is achieved, initially, through methylation (Fig. 1*c*). An understanding of the chemical logic of cobalamin synthesis revolves around the modifications that take place on C20, and this paper directly addresses this issue through the study of the anaerobic pathway C20 methyltransferase, CbiL.

* The research was supported by Grants BBS/B04145 and BB/E002137 from the Biotechnology and Biological Sciences Research Council. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

This work is dedicated to the memory of Prof. Ian Scott.

§ The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1 and S2.

The atomic coordinates and structure factors (code 2QBU) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ To whom correspondence may be addressed. Tel.: 44-1227-824690; E-mail: m.j.warren@kent.ac.uk.

² To whom correspondence may be addressed. Tel.: 44-20-7882-6360; E-mail: r.w.pickersgill@qmul.ac.uk.

³ The abbreviations used are: SAM, *S*-adenosyl-L-methionine; SAH, *S*-adenosyl-L-homocysteine; SUMT, *S*-adenosyl-L-methionine uroporphyrinogen III methyltransferase; MES, 4-morpholineethanesulfonic acid; ITC, isothermal titration calorimetry; ALA, 5-aminolevulinic acid.

Structure and Function of CbiL

CbiL was first identified as the C20 methyltransferase when the *Salmonella enterica* enzyme was reported to be able to methylate precorrin-2, albeit in very low yields (9). Sometime later, the *S. enterica* enzyme was shown to have a preference for metal-containing substrates (10). The enzyme was shown to be able to methylate both cobalt-precorrin-2 and cobalt factor II, as well as zinc derivatives. Moreover, the enzyme does not apparently discriminate on the basis of the oxidation state of the metal ion and was found to methylate both Co(II) and Co(III) forms. However, no rates were given for the reaction, and thus no substrate preference was identified. Therefore, there remains some dispute as to the true identity of the pathway intermediate (2, 11). Very recently, a structure of an enzyme purported to be CbiL, based solely on sequence similarity and with no supporting functional data, has been reported (12).

Sequence analysis has revealed that many of the enzymes of the aerobic and anaerobic pathways are similar, suggesting that the two pathways have a common evolutionary origin (3, 6). This is most clearly seen with many of the methyltransferases, where aerobic and anaerobic homologues exist for the enzymes that methylate at C20 (CobI and CbiL) (Fig. 1c), C17 (CobJ and CbiH), C11 (CobM and CbiF), and C5 (CobL and CbiE) (6). Moreover, the majority of the methyltransferases associated with cobalamin biosynthesis share sequence similarity, indicating that they are all derived from a common ancestor (6, 7). A study of these methyltransferases will not only provide some significant insight into how regio-specificity can be acquired within a common protein framework but also provide molecular detail on how these enzymes discriminate between metal-free and cobalt-containing substrates, which is one of the major differences between the two pathways.

In this paper, we look at the structure and function of the *Methano-*

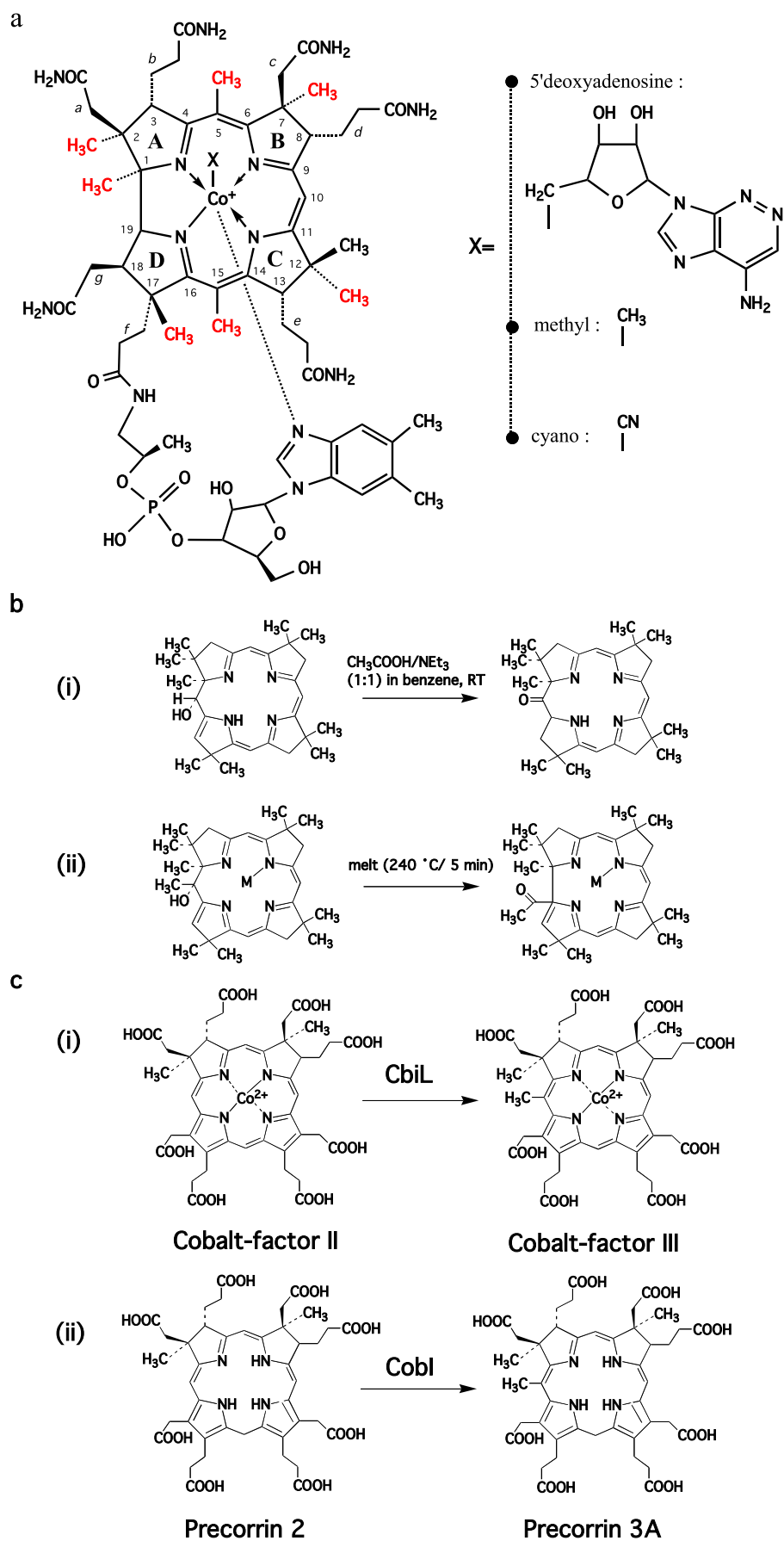


TABLE 1
Bacterial strains*B.m.*, *B. megaterium*, *Mth*, *M. thermautotrophicus*; *S.e.*, *S. enterica*, *P.d.*, *P. denitrificans*.

Strains or plasmid	Genotype and/or phenotype	Description	Reference or source
AR3711	<i>S.e. cbiL</i> mutant, <i>lacI</i> ^P		Ref. 14
pAR8547	<i>S.e. cbiL</i>	DdeI 10583-PstI 11432 fragment cloned into pKK223.3	Ref. 14
pET14b <i>Mth cbiL</i>	<i>Mth cbiL</i>	Cloned by PCR NdeI and BamHI into pET14b	This study
pET3a <i>Mth cbiX</i>	<i>Mth cbiX</i>	Cloned by PCR NdeI and BamHI into pET3a	Ref. 33
pER242	<i>P.d. cobA</i>	Cloned by PCR NdeI and BamHI into pET14b	Ref. 34
pET14b <i>Mth hemB</i>	<i>Mth HemB</i>	Cloned by PCR NdeI and BamHI into pET14b	This study
pSD1	<i>B.m. hemC</i>	Cloned by PCR NdeI and BamHI into pET14b	Ref. 13
pSD2	<i>B.m. hemD</i>	Cloned by PCR NdeI and BamHI into pET14b	Ref. 13
pHL103	<i>B.m. sirC</i>	Cloned by PCR NdeI and BamHI into pET14b	Ref. 13
pSF23	<i>Mth cbiL</i> ^{D104A}	Point mutation at bp 311 from GAT to GCT, aspartate 104 to alanine, template pET14b <i>Mth cbiL</i>	This study
pSF24	<i>Mth cbiL</i> ^{K176A}	Point mutations at bp 526/527 from AAG to GCG, lysine 176 to alanine, template pET14b <i>Mth cbiL</i>	This study
pSF25	<i>Mth cbiL</i> ^{Y220A}	Point mutations at bp 658/659 from TAT to GCT, tyrosine 220 to alanine, template pET14b <i>Mth cbiL</i>	This study
pSF26	<i>Mth cbiL</i>	NdeI-BamHI fragment of pET14b <i>Mth cbiL</i> cloned into NdeI-BamHI of pETac	This study
pSF27	<i>Mth cbiL</i> ^{D104A}	NdeI-BamHI fragment of pSF23 cloned into NdeI-BamHI of pETac	This study
pSF28	<i>Mth cbiL</i> ^{K176A}	NdeI-BamHI fragment of pSF24 cloned into NdeI-BamHI of pETac	This study
pSF29	<i>Mth cbiL</i> ^{Y220A}	NdeI-BamHI fragment of pSF25 cloned into NdeI-BamHI of pETac	This study
pETcoco-2 ABCDC	<i>M. bar cobA-Mth hemB-B.m. hemC-D-Mth sirC</i>	Genes cloned in order in pETcoco-2	This study

thermobacter thermautotrophicus CbiL. We determine the nature of the true substrate for the enzyme and elucidate the structure of CbiL. From this we have been able to provide experimental evidence for how the substrates bind to the enzyme, how it recognizes cobalt-containing substrates, and how it forms a productive ternary complex.

EXPERIMENTAL PROCEDURES

Chemicals and Reagents—Most chemicals, DEAE-Sephacel, and antibiotics were purchased from Sigma unless otherwise stated. Bacterial strains were purchased from Novagen, Invitrogen, or Promega. Strains and plasmids used in this work are described in Table 1.

Cloning, Overexpression, and Purification—*M. thermautotrophicus cbiL* was amplified via PCR and cloned using the pGEM[®]-T Easy vector system from Promega. The gene was subsequently subcloned into the NdeI and BamHI sites of the pET14b and pETac vectors. For *in vivo* activity analysis, CbiL was subcloned into pETac, a vector that is similar to pET14b but instead of the T7 promoter contains a *lac* promoter. The CbiL mutants D104A, K176A, and Y220A were generated using the QuikChange[®] II site-directed mutagenesis kit (Stratagene). For protein overproduction, *Escherichia coli* strain Rosetta(DE3)pLysS was transformed with pET14b *Mth cbiL*. The recombinant strain was grown in super Luria-Bertani broth with ampicillin and chloramphenicol at 37 °C and shaken vigorously to an A₆₀₀ of ~0.6. Protein expression was induced with 0.4 mM isopropyl 1-thio-β-D-galactopyranoside overnight at 16 °C. The cells were collected by centrifugation at 4,000 rpm for 15 min at 4 °C. The pellet was resuspended in 10 ml of 20 mM Tris-HCl buffer, pH 8.0, containing 0.5 M NaCl and 5 mM imidazole. Cells were lysed, and cell debris was removed by centrifugation. CbiL was purified by immobilized metal ion affinity

chromatography. The bound protein was isolated by competitive elution with increasing concentrations of imidazole and finally eluted in 10 ml of 20 mM Tris-HCl buffer, pH 8.0, containing 0.5 M NaCl and 400 mM imidazole.

Substrate Synthesis—CbiL substrates were synthesized by multienzyme synthesis (13) under nitrogen containing less than 3 ppm oxygen in an anaerobic glove box (Belle Technology). *E. coli* constructs overexpressing the genes 5-aminolevulinic acid (ALA) dehydratase (*M. thermautotrophicus hemB*), porphobilinogen deaminase (*Bacillus megaterium hemC*), uroporphyrinogen III synthase (*B. megaterium hemD*), SUMT (*Pseudomonas denitrificans cobA*), precorrin-2 dehydrogenase (*B. megaterium sirC*), and cobalt chelatase (*M. thermautotrophicus cbiX*) were overproduced separately and lysed. For factor II production, cell-free lysates of HemB (10 mg), HemC, HemD, CobA, and SirC (5 mg) were pooled with 20 mg of SAM, 13 mg of NAD⁺, and 10 mg of the starter molecule ALA in a total volume of 40 ml in 50 mM Tris-HCl buffer, pH 8.0. Oxygen was removed by repeatedly freeze-thawing the reaction mix in liquid nitrogen under a strong vacuum. The sample was transferred into the glove box and incubated overnight at room temperature. The reaction mixture was filtered (0.2 μm; Millipore) prior to incubation with affinity-purified CbiX (7.5 mg) and CoCl₂ (200 μM).

Alternatively, to obtain highly concentrated cobalt factor II for EPR analysis and crystal soaking experiments, a single plasmid, pETcoco-2ABCDC, carrying the genes *M. thermautotrophicus hemB* and *sirC*, *B. megaterium hemC* and *hemD*, and *Methanosarcina barkeri cobA* was overexpressed. Lysate of the *E. coli* strain carrying the plasmid was purified (nickel-chelating Sepharose), and the most concentrated protein fraction was transferred into the glove box. Following buffer exchange of the

FIGURE 1. *a*, structure of cobalamin, with the peripherally added methyl groups highlighted in red. Although eight SAM-derived methyl groups are added, only seven are observed in the final corrin ring, since the methyl group added to C20 is lost during the ring contraction process. *b*, chemical studies demonstrating that whereas a secondary alcohol at position C20 generates a 20-oxotetrahydrocorphin derivative (*i*), the tertiary alcohol can undergo ring contraction (*ii*). *c*, the reactions catalyzed by CbiL (*i*) and Cobl (*ii*).

TABLE 2

X-ray data and refinement statistics

These data were collected at ID14-3, wavelength 0.931 Å, and the space group was P6₂.

Crystal	Native	MeHgCl	SmCl ₃
Soak concentration	None	1 mM	50 mM
Soak time (h)	None	0.5	3
Cell edges: <i>a</i> = <i>b</i> , <i>c</i> (Å)	135.92, 50.37	135.74, 50.26	135.99, 50.29
Resolution range (Å) ^a	50.38 to 2.21 (2.21 to 2.10)	58.82 to 2.33 (2.45 to 2.33)	58.93 to 2.50 (2.64 to 2.5)
<i>I</i> / <i>σ</i> (<i>I</i>) ^a	18.6 (4.8)	18.5 (3.3)	24.2 (6.9)
Multiplicity ^a	7.5 (7.2)	7.0 (3.6)	9.1 (7.2)
Completeness (%) ^a	100 (100)	97.1 (80.3)	100 (100)
Midslope anomalous probability		1.326	1.245
Phasing power (acentric)		1.18	1.04
<i>R</i> _{cutis} (acentric)		0.82	0.85
<i>R</i> _{merge} ^a	0.074 (0.363)	0.076 (0.308)	0.068 (0.249)
Root mean square bond distance (Å) ^a Average <i>σ</i>	0.017 (0.022)		
Root mean square angle (degrees) ^a Average <i>σ</i>	1.753 (1.993)		
R-factor	19.2		
Free R-factor	27.0		
Average B-factors (Å ²)			
Main chain	32.7		
Side chains	38.1		
All atoms	35.6		
Validation PROCHECK	438 waters 3980 total atoms in file		
G-factor	−0.03		
Ramachandran plot (favored/allowed/generously allowed) (%)	93.5/6.3/0.3		

^a Values in parentheses are for the outer shell.

protein into 50 mM Tris-HCl buffer, pH 8.0, 3 ml of the protein was pooled with 2 ml of cofactor mix consisting of 20 mg/ml SAM, 6.5 mg/ml NAD, and 10 mg/ml ALA in 50 mM Tris-HCl buffer, pH 8.0. The cofactor mix was adjusted to pH 8.0 with 2 M NaOH prior to use. The reaction was incubated overnight at room temperature. Cobalt factor II was produced as described above by the addition of *M. thermotrophicus* CbiX and CoCl₂, depending on the factor II concentration. Cobalt factor II was DEAE-purified prior to use.

In Vitro Activity Assay—The C20 methyltransferase assay used cobalt factor II (5.6 μM), SAM (950 μM), and varying amounts of wild type CbiL and mutant variants. The disappearance of cobalt factor II at λ_{max} 386 nm was monitored using a Hewlett Packard 8452A photodiode array spectrophotometer. The concentration of cobalt factor II was determined using an extinction coefficient of 1.6 × 10⁵ M^{−1} cm^{−1}.

In Vivo Activity Assay—An *S. enterica* *cbiL* mutant strain (AR3711) was transformed with a plasmid containing wild type *cbiL* (pSF26) and the three *cbiL* mutant plasmids pSF27, pSF28, and pSF29. The negative control contained pETac, the positive control *S. enterica* *cbiL* (pAR8547). Transformations were spread on LB agar plates containing the appropriate antibiotic, incubated for 24 h at 37 °C. A few colonies were picked from each plate and transferred to agar analytic plates (14). If the *S. enterica* *cbiL* mutant strain reacquired the ability to produce cobalamin, the colony developed a red color.

Crystallization of CbiL—After purification, CbiL was desalted using Sephadex 25, and the His tag was cleaved by incubating overnight with thrombin at 4 °C. The protein was

subsequently gel-filtered (SuperdexTM 200 HR 10/30; Amersham Biosciences) into 50 mM Tris-HCl buffer, pH 8.0, containing 100 mM NaCl before concentration to 2.5 mg ml^{−1} for crystallization. Hanging drop vapor equilibration trials using reservoirs containing 100 mM MES buffer, pH 6.5, and 2.5 M ammonium sulfate and protein drops comprising equal volumes of CbiL with 3.3 mM SAM and reservoir produced clusters of needle crystals.

Two heavy atom derivatives were obtained by soaking native CbiL crystals in reservoir supplemented with mercury chloride (MeHgCl) and samarium chloride (SmCl₃), respectively (see Table 2 for soak details). The diffraction quality was not affected by the inclusion of cryoprotectant, 10% (v/v) glycerol, in the soak solutions, and previtrified crystals were transported to the synchrotron. Native data were initially collected using PX 9.6 at the Synchrotron Radiation Source (Daresbury, UK). The two heavy metal derivative data sets and high resolution native data were collected at station ID 14-3 at the European Synchrotron Radiation Facility (Table 2).

Substrate Soaking into Native CbiL Crystals—Cobalt factor II was prepared as described above. The solution was passed through a cobalt-loaded ion affinity chromatography resin to remove the biosynthetic enzymes and eluted in 6 ml of 50 mM Tris-HCl buffer, pH 8.0, with a concentration of 170 μM. After freeze-drying, the substrate was resuspended in mother liquor consisting of 100 mM MES buffer, pH 6.5, containing 2.47 M ammonium sulfate and 10% glycerol (v/v). A drop (8 μl) of substrate was placed on the same coverslip as the target crystal, and the crystal was transferred into the substrate solution. The

coverslip was quickly placed back onto the well to control the crystallization of ammonium sulfate. The crystals were soaked for 10 min in close to saturating concentrations of tetrapyrrole substrate and then harvested onto a loop and vitrified immediately in liquid nitrogen. All steps, including crystal growth, substrate preparation, crystal soaks, and cryocooling, were under anaerobic conditions. Data were collected from seven tetrapyrrole-soaked crystals, some of which became colored upon soaking. The majority of the crystals diffracted to around 2.0 Å resolution, the best 1.8 and worst 3.0 Å. Additional density could be detected in σ_A -weighted $2F_o - F_c$ Fourier syntheses for three of the seven soaks, although the density was low, reflecting low occupancy and multiple binding modes. Cobalt factor II was translated and rotated into the electron density present in the initial maps in what appears to be the principal binding mode in the crystal.

Structure Determination and Refinement—Diffraction data were processed using MOSLEM (15), SCALA, and the CCP4 program suite (16). MOLREP (17) and the coordinates for CbiF were used for molecular replacement calculations, MLPHARE (18) was used for heavy atom phasing, and density modification was by DM (19), which included averaging the two copies of the CbiL subunit in the asymmetric unit along with the usual histogram matching and solvent flattening. Fourteen cycles of refinement and rebuilding using REFMAC5 (20) and "O" (21), respectively, gave the final structure. The structure was validated using PROCHECK (22).

HPLC-Mass Spectrometry—Mass spectroscopic data were obtained on an Agilent 1100 liquid chromatography system connected to an Agilent 1100 liquid chromatography/MSD Trap, using the electrospray ionization technique in the positive mode. Spectra were monitored by DAD-UV detection. Samples of cobalt factor II and cobalt factor III were prepared in the glove box, filtered (0.2 μm ; Millipore), and transferred into HPLC vials with air-tight seals. Routinely, 100 μl of sample was injected into a BDS HYPERSIL C18 column (dimensions, 250 \times 4.6 mm; particle size, 5 μm) and eluted with a gradient of 50% of 1 M ammonium acetate containing 9% acetonitrile and 50% double-distilled H₂O increasing within 5 min to 100% of 1 M ammonium acetate containing 9% acetonitrile at a flow rate of 1 ml/min.

EPR—EPR samples of cobalt factor II and cobalt factor III were prepared using the multienzyme approach in 50 mM Tris-HCl buffer, pH 8.0, containing 100 mM NaCl. The compounds were DEAE-purified. Aerobic samples of these two compounds were prepared anaerobically but removed from the glove box and exposed to air before freezing in liquid nitrogen. EPR samples of cobalt factor II incubated with *S*-adenosyl-L-homocysteine (SAH) and CbiL or SAH and the mutants D104A, K176A, and Y220A were also prepared. Factor II was prepared in 50 mM HEPES buffer, pH 8.0, containing 100 mM NaCl using the single plasmid method. Subsequently, cobalt factor II was produced, purified, and concentrated on a DEAE column. CbiL and the three CbiL mutants were purified, the His tag was cleaved, and they were transferred into the glove box and buffer-exchanged into 50 mM HEPES buffer, pH 8.0, containing 100 mM NaCl. EPR samples were made anaerobically in 50 mM HEPES buffer, pH 8.0, with final concentrations of 60 μM for cobalt factor II

and CbiL, respectively, and 780 μM SAH. EPR spectra were obtained at X-band using a Bruker ELEXSYS E500 spectrometer, equipped with an Oxford Instruments ESR900 liquid helium cryostat. The run conditions for the EPR measurements were as follows: temperature 20 K, microwave power 0.5 milliwatts, modulation amplitude 0.5 milliteslas.

Isothermal Titration Calorimetry (ITC)—The thermodynamics of SAM and SAH binding to CbiL were determined using an isothermal titration microcalorimeter (VP-ITC MicroCal™). SAM or SAH was dissolved in 50 mM Tris-HCl buffer, pH 8.0, containing 100 mM NaCl. Prior to the experiment, samples were degassed under vacuum for 5 min. The sample cell was filled with the protein (200 μM). Initially, titrations with the cofactors consisted of a one-time injection of 2 and 5 μl for SAM and SAH, respectively. This was followed by 59 5- μl and 29 10- μl injections for SAM and SAH, respectively. All runs were at 30 °C. Heat release was monitored for each injection peak and analyzed using the Microcal Origin software, a data analysis and graphics software. The integrated heat was plotted against the molar ratio of SAM or SAH added to CbiL in the cell. A complete binding isotherm for the interaction was obtained. The one-site model was used to fit the data as implemented by the Microcal Origin software based on the Wiseman isotherm (23). The values for stoichiometry, dissociation constant (K_D), and enthalpy (ΔH), were determined.

RESULTS

The *M. thermautotrophicus* cbiL Encodes a Cobalt Factor II Methyltransferase—The *M. thermautotrophicus* cbiL was amplified by PCR from genomic *M. thermautotrophicus* DNA (24), and the 714-bp product was cloned into pET14b, thereby allowing the encoded protein to be produced with an N-terminal histidine-rich peptide extension. The fidelity of the cloning procedure was confirmed by sequencing.

In order to probe the physiological activity of the His-tagged CbiL, the fused gene was subcloned into the plasmid pET-ac, which is a pET plasmid derivative under the control of a *tac* promoter (instead of the T7 promoter). The *tac* promoter allows expression in strains that do not contain a T7 RNA polymerase. The *S. enterica* cbiL mutant strain (AR3711) was transformed with pETac-cbiL, pETac (negative control) and pAR8547 (positive control, which harbors the *S. enterica* cbiL). Colonies were restreaked on minimal medium plates containing propanediol and a pH-sensitive indicator (14). If cobalamin is made by the bacteria, they are able to convert the propanediol to propionic acid using their B₁₂-dependent diol dehydratase. The production of the acid turns the indicator red and results in the appearance of red colonies on the plate. Both pAR8547 and pETac-cbiL were found to complement the *S. enterica* cbiL mutant and generate red colonies, whereas the negative control (pETac) colonies remained white, indicating that they were unable to produce cobalamin. This result demonstrates not only that the function ascribed to *M. thermautotrophicus* cbiL was correct but moreover that the His-tagged *M. thermautotrophicus* cbiL is also active *in vivo*.

When pET14b-cbiL was transformed into the *E. coli* strain BL21(DE3)pLysS, it resulted in the production of recombinant CbiL as adjudged by the appearance of a 29-kDa band as

Structure and Function of CbiL

assessed by SDS-PAGE of crude cell extracts of the strains (Fig. S1a). The overproduced protein was purified by metal affinity chromatography. From 1 liter of culture, ~25 mg of pure protein were reproducibly obtained. The purified protein ran as a single band on a SDS-PAGE with an apparent molecular mass of 29 kDa, in close agreement with the expected molecular mass of 26.2 kDa plus ~2.2 kDa corresponding to the N-terminal His tag.

The native state of the protein was analyzed by analytical gel filtration chromatography on a Superdex 200 column. The protein eluted from the column with a retention time that corresponded to a molecular mass of 51 kDa, indicating that the enzyme exists as a homodimer (data not shown). CbiL was also analyzed by nondenaturing gel electrophoresis where the protein migrated as a single tight band, consistent with the size of a dimer.

The activity of *M. thermautotrophicus* CbiL as a cobalt factor II methyltransferase, was monitored *in vitro*, utilizing a simple anaerobic *in vitro* assay involving the incubation of cobalt factor II with SAM and CbiL. During the incubation, a color change was observed from dark green, the color of cobalt factor II, to a novel bright green intermediate, which was identified as cobalt factor III by mass spectrometry with the addition of 14 mass units. The change in color was associated with a minor change in the absorption spectrum, where the main Soret peak of cobalt factor II at 386 nm shifted to 392 nm (Fig. S1b). This change in the electronic spectrum was exploited to develop a quantitative assay to monitor the enzymatic reaction. Using purified CbiL, a specific activity of 2.0 ± 0.3 nmol/min/mg was calculated. Since CbiL is derived from a thermophilic archaea that grows optimally at 65 °C, the enzyme was expected to be relatively heat-stable and more active at higher temperatures. Therefore, the assay was repeated at 37 and at 65 °C, where specific activities of 4.1 ± 0.8 and 13.6 ± 2.4 nmol/min/mg were observed, respectively.

Product of the CbiL Reaction Is Co(II) Factor III—In the oxygen independent cobalamin pathway, the cobalt ion bound by the tetrapyrrole ring has been proposed to play a redox role during the ring contraction process, where a Co(III) to Co(I) transition has been implicated to help mediate lactone ring formation (25). If this proposal is correct, then the trimethylated product of the reaction catalyzed by CbiL should be Co(III). Therefore, it is important to record the oxidation states of the substrate cobalt factor II and the new intermediate cobalt factor III and determine whether *M. thermautotrophicus* CbiL has preferences for the redox state of the cobalt ion incorporated into its substrate. To investigate this question, EPR spectrometry was used as a tool to monitor the oxidation state of the metal during the enzymatic transformation.

The EPR spectra obtained from the aerobic and anaerobic cobalt factor II and cobalt factor III samples over a broad field range are shown in Fig. 2. The spectrum of Co(II) factor II shows an axial spectrum with extensive hyperfine splitting known to arise from low spin Co(II). The action of CbiL, under anaerobic conditions, alters the spectrum slightly due to the methylation of Co(II) factor II to Co(II) factor III. However, the basic axial spectrum of a low spin Co(II) species with a *g* perpendicular value of 2.52 is maintained. When exposed to air, the

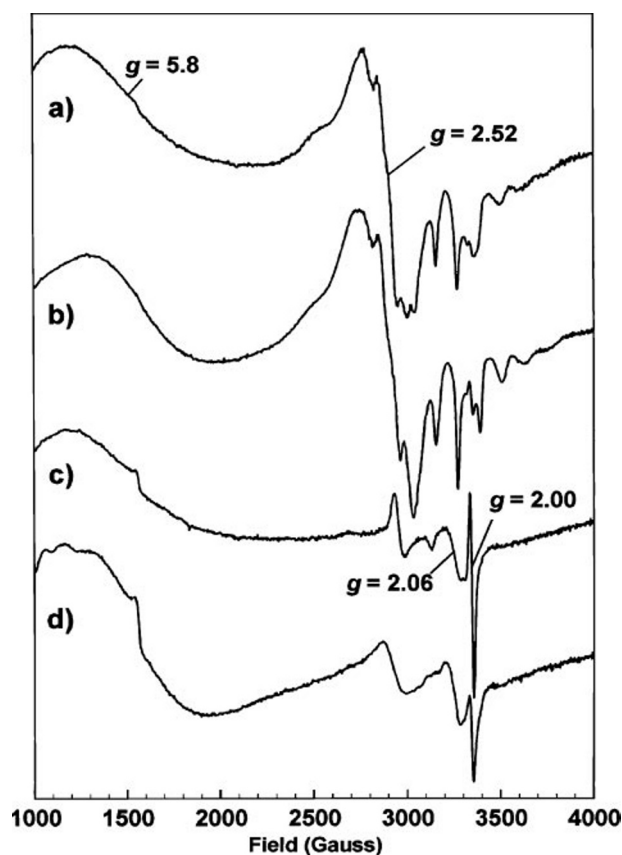


FIGURE 2. EPR spectroscopy at X-band showing Co(II) factor II (a), Co(II) factor III (i.e. following the action of CbiL) (b), Co(III) factor II (c), and Co(III) factor III (i.e. following the action of CbiL) (d). The significance of the marked *g* values is explained under "Results."

same samples show no or very little low spin Co(II) signal. Therefore, the cobalt has been largely oxidized to Co(III). Some residual signal at *g* = 2.06 arises from contaminating Cu(II), probably in the EPR cavity. The broad signals at *g* = 5.8 arise from high spin Co(II), attributed to the hydrated Co(II) ion. These signals are not greatly diminished in the presence of air, indicating that rapid oxidation is a property of the low spin, tetrapyrrole-bound, Co(II).

Substrate Specificity—The discussion of the redox states of intermediates also raises the question of substrate specificity of *M. thermautotrophicus* CbiL. Thus, the enzyme was incubated separately with Co(III) factor II and Co(II)-precorrin-2. The enzyme was found to methylate Co(III) factor II, although the conversion of Co(II)-precorrin-2 with CbiL was much slower and resulted in a ratio of 60% cobalt-precorrin-3 to 40% cobalt-precorrin-2 in the sample at the end of the incubation. Therefore, it can be concluded that *M. thermautotrophicus* CbiL is first a cobalt factor II methyltransferase with less ability to catalyze the methyltransfer onto the more reduced dipyrrocorphin species. The metal-free intermediate, factor II, however, did not act at all as a substrate for CbiL. Thus, cobalt or other metals (10) have to be inserted into the tetrapyrrole ring for the C-20 methylation to take place.

Crystallization of CbiL—Purified CbiL was concentrated and desalted, and its N-terminal histidine-rich tag was removed by cleavage with thrombin. The cleaved protein was concentrated

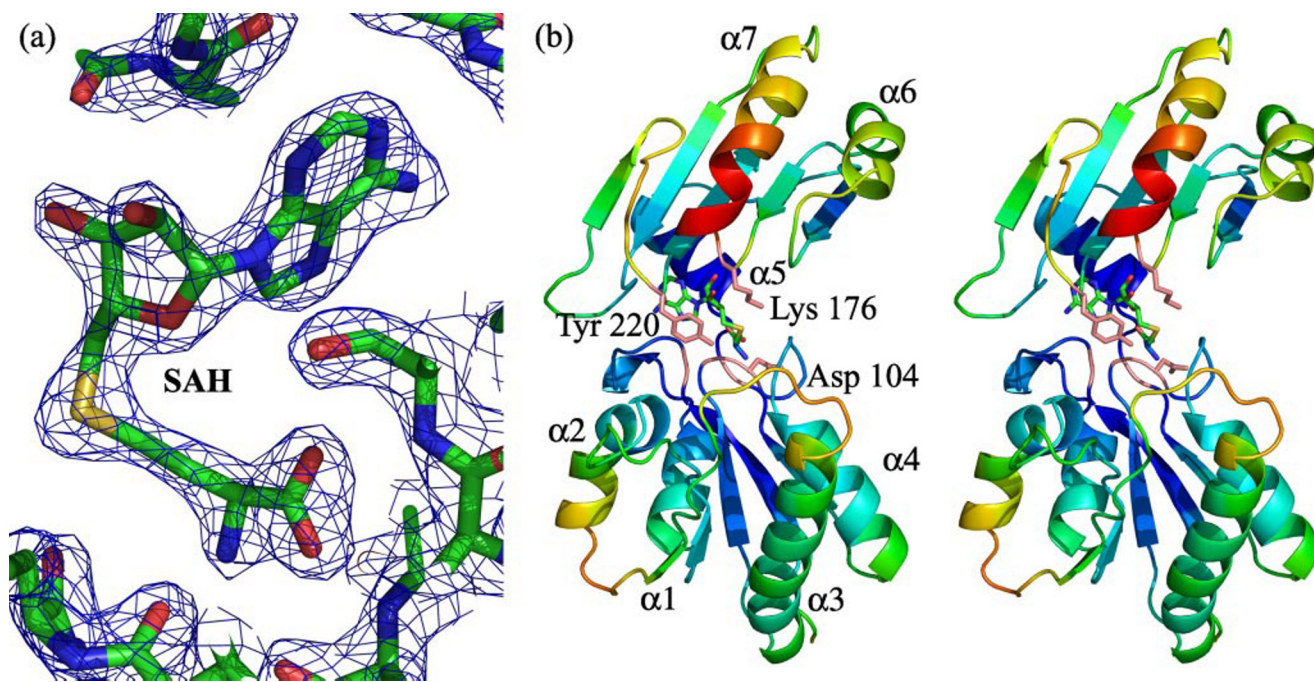


FIGURE 3. *a*, σ_A -weighted $2F_o - F_c$ Fourier synthesis showing the 2.1 Å electron density for SAH. The electron density is represented as the blue mesh, contoured at 1.5σ , and SAH is shown in a stick representation. *b*, stereo view of the architecture of *M. thermautotrophicus* CbiL (schematic diagram representation colored according to temperature factor, red (hot) and blue (cold)) with SAH shown in atom-colored stick representation and the residues conserved across the CbiLs shown in salmon. Invariant glycines (positions 11, 103, and 130) are painted on the schematic diagram, and the side chains of Asp¹⁰⁴, Lys¹⁷⁶, and Tyr²²⁰ are shown in a stick representation. All of the invariant residues are close the SAH binding site, the domain, the interface, the dimer interface, and the tetrapyrrole binding cleft. The loop connecting β_2 to extended α_3 has relatively high temperature factors and differs in conformation between subunits A and B due to the different environments in the crystal.

and purified further by gel filtration chromatography. The fast protein liquid chromatography-purified CbiL fractions were combined, concentrated, and entered into crystallization trials, which were performed at 18 °C via hanging drops using the vapor diffusion technique. Observation of microcrystalline precipitate led to refinement of conditions (100 mM MES buffer, pH 6.5, containing 2.5 M ammonium sulfate) that produced diffraction quality crystals. Crystals grew only in the presence of exogenously added SAM or SAH.

The structure of CbiL was solved using a combination of molecular and isomorphous replacement as well as noncrystallographic symmetry averaging. CbiF (Protein Data Bank code 1cbf) (26), which has 28% sequence identity with CbiL, was used as a search model in molecular replacement. The volume of the unit cell and the self-rotation function were consistent with one dimer in the asymmetric unit, and the dimer was therefore used as the search model. The pattern of absences was consistent with space group $P6_2$ or $P6_4$, and the molecular replacement solution supported the former as the true space group, with correlation coefficients 0.213 and 0.173, respectively. The molecular replacement solution would not refine further, and the quality of the 2.5 Å σ_A -weighted $2F_o - F_c$ map was poor; however, the phases were successfully used to find two mercury and two samarium sites in difference Fourier syntheses. These heavy atom sites were refined, and protein phases were improved using density modification and refined as described under "Experimental Procedures." Fourteen rounds of rebuilding and refinement produced the final model, which has an R-factor and R-free of 19.2 and 27.0% at 2.1 Å. The electron density for SAH is shown in Fig. 3*a*.

Structure of CbiL—We determined the structure of *M. thermautotrophicus* CbiL at 2.1 Å. The model comprises all 228 residues for both subunits comprising the biologically authentic dimer present in the asymmetric unit. There is one SAH molecule bound per subunit (Fig. 3*b*). The subunit comprises two topologically distinct domains with a single polypeptide connection between them. The N-terminal domain has a central five-stranded parallel β -sheet, and the C-terminal domain has a central five-stranded mixed β -sheet. The β -sheets are conserved across the five known structures, although the β -strands do vary in length. The loops connecting the strands may contain helices, and it is the presence or absence of helices and their position and orientation that varies most across the structures. Comparing the C-terminal domain of *M. thermautotrophicus* CbiL with the recently published *Chlorobium tepidum* CbiL structure (Protein Data Bank code 2E0N) (12), *M. thermautotrophicus* CbiL has 2 α helices above the sheet and 1 α helix below compared with *C. tepidum* CbiL, with 3 α helices above and 2 α helices below (Figs. 4*a* and 5*c*). The differences are not restricted to the C-terminal domain, since the *M. thermautotrophicus* CbiL has an extended α_3 and shortened β_2 and β_4 compared with the *C. tepidum* structure.

In comparison with all of the known structures of the class III methyltransferases (27) (SUMT, Protein Data Bank code 1S4D; CbiF, 1CBF; CbiE, 2BB3) (28, 29), the conformation of the SAH molecules bound between the two domains of the subunit is closely similar. In fact, superimposition of the SAH molecules is the most convenient way of superimposing these protein structures (Fig. 4, *a–d*). The *C. tepidum* CbiL and the *B. megaterium*

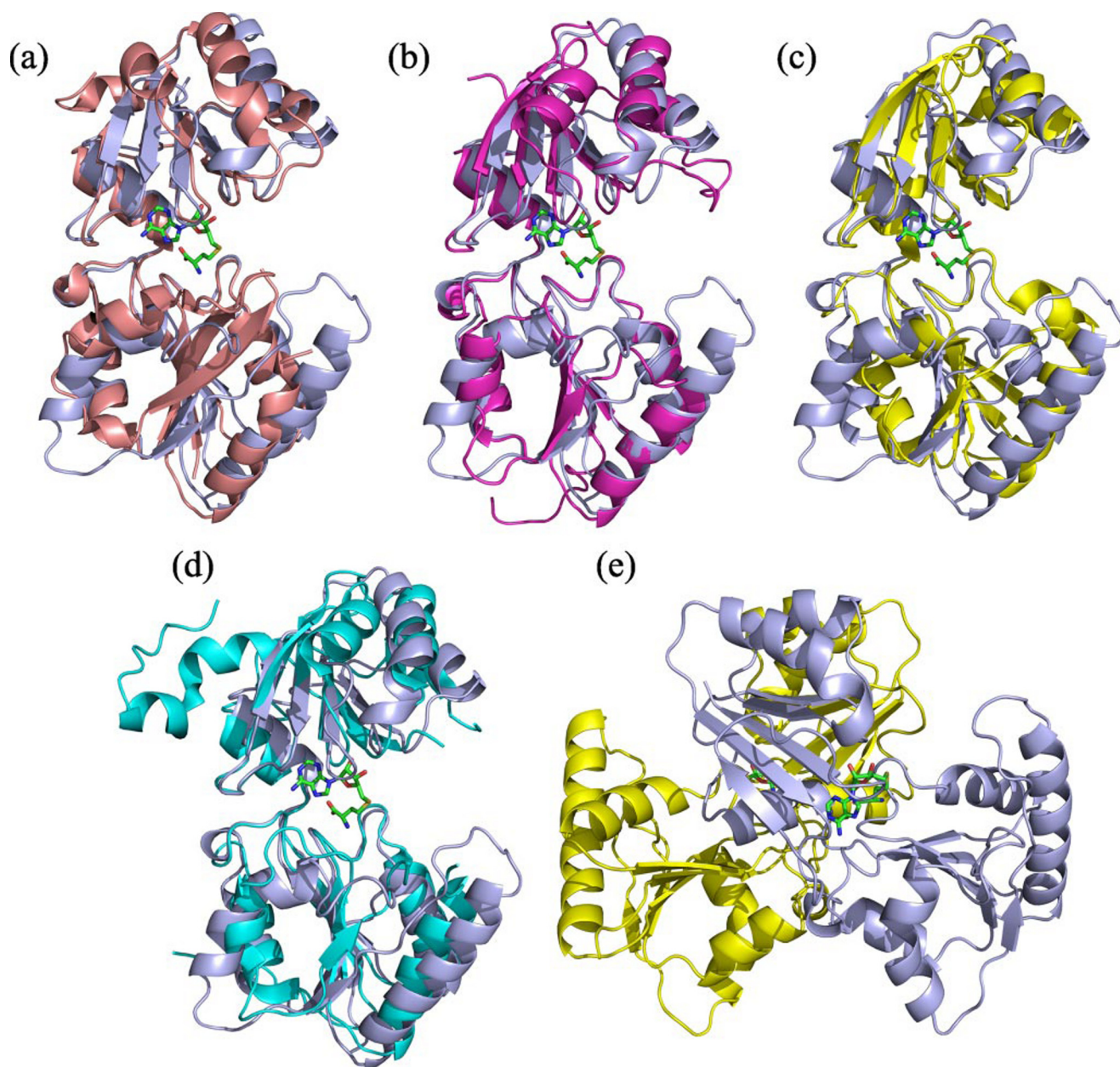


FIGURE 4. **Structure of CbiL.** *a–d*, the methyltransferases *C. tepidum* CbiL (Protein Data Bank code 2E0N), CbiF (1CBF), CobA (1S4D), and CbiE (2BB3) each superimposed in turn on *M. thermoautotrophicus* CbiL. The superimposition was best achieved by superimposing the SAH molecules present in the structures. Considerable diversity in the number, length, and position of the helices is seen in these structures. *e*, the *M. thermoautotrophicus* CbiL dimer with SAH in a stick representation. The asymmetry in the β 2-loop- α 3 connection can be seen at the extreme left and right edges.

CbiF are the closest structures to *M. thermoautotrophicus* CbiL (32% sequence identity, root mean square deviation 2.0 Å over 855 atoms and 28% identity, 2.5 Å root mean square deviation over 1000 equivalenced atoms, respectively) (12, 26). The *Archaeoglobus fulgidus* CbiE is the most different with a root mean square deviation of 5.5 Å for 794 equivalenced atoms.

Biochemical evidence and the crystal structure clearly indicate that CbiL is a homodimer, which crystallized with one dimer in the asymmetric unit (Fig. 4e). The five C-terminal β -strands of each monomer lead to 10 β -strands in the homodimer. A similar arrangement of β -strands has been seen across the class III methyltransferase family.

Conserved Residues in C20 Methyltransferases—Only six residues are invariant in all C-20 methyltransferases. These are

Gly¹¹, Gly¹⁰³, Asp¹⁰⁴, Gly¹³⁰, Lys¹⁷⁶, and Tyr²²⁰ as highlighted in Fig. 3b. Gly¹¹, Gly¹⁰³, and Asp¹⁰⁴ are involved in SAM binding via formation of a nucleotide binding pocket. Gly¹¹ is part of the typical GXGXXG SAM binding region conserved throughout the class III methyltransferases. One residue, Gly¹³⁰, is located in the linker region β 5- α F between the N- and C-terminal domains. Gly¹¹, Gly¹⁰³, Asp¹⁰⁴, and Gly¹³⁰ are not only conserved in CobIs and CbiLs but also in a variety of other methyltransferases, such as CbiF, CysG, CobA, and CbiE. Significantly, the two other conserved residues, Lys¹⁷⁶ and Tyr²²⁰, are restricted to the C20 methyltransferases. In *P. denitrificans* SUMT and *S. enterica* CysG, Lys¹⁷⁶ is replaced by methionine and in *B. megaterium* CbiF by leucine. Met¹⁸⁴ (equivalent to Lys¹⁷⁶) from *P. denitrificans* SUMT makes van der Waals con-

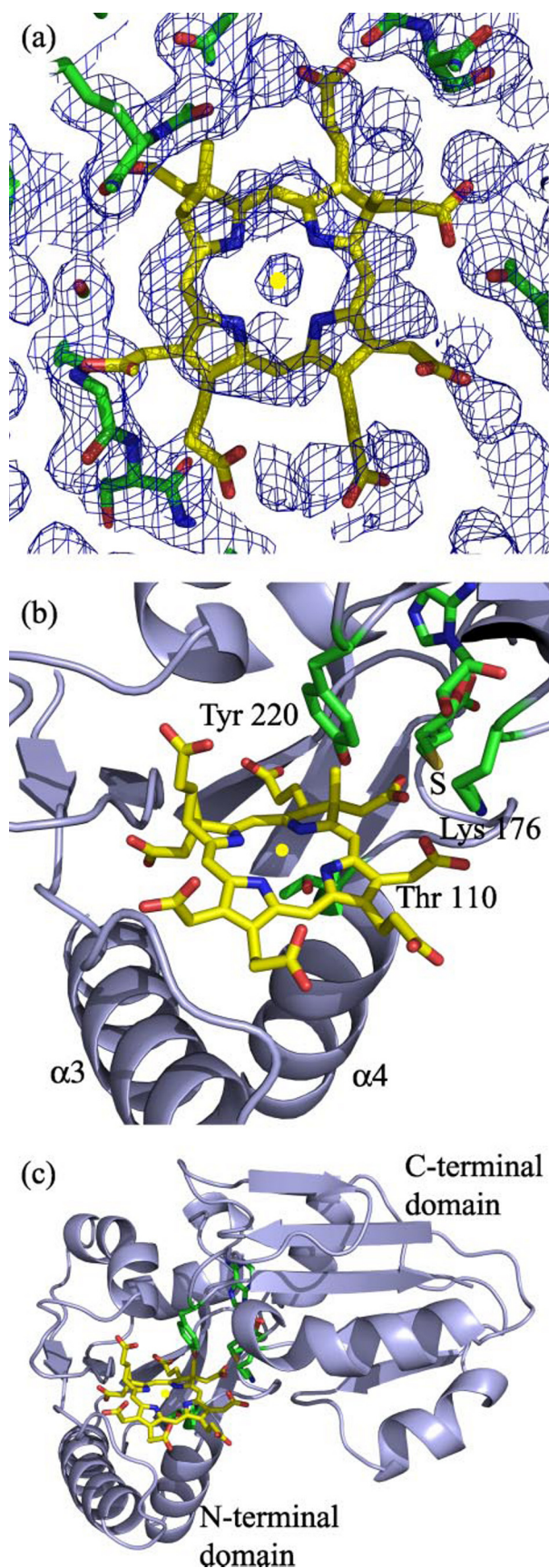


FIGURE 5. Tetrapyrrole binding to *M. thermautotrophicus* CbiL. *a*, σ_A -weighted $2F_o - F_c$ Fourier synthesis showing the 2.1 Å electron density for the tetrapyrrole contoured at 0.5σ . The occupancy is low, there is no density for

tact with SAM, whereas Met from *S. enterica* CysG does not contact SAM but is thought to be involved in SAM binding by keeping order in the active site residues (28). In *M. thermautotrophicus* CbiL, Lys¹⁷⁶ is located in close proximity to SAH but does not make contact. Since the residue is also close to the proposed tetrapyrrole binding region, it is well positioned to interact with the acetic and propionic acid side chains of the macrocycle.

The S-Adenosyl-L-homocysteine Binding Site—Although *M. thermautotrophicus* CbiL was crystallized in the presence of SAM, it is actually SAH that is found bound in the structure, because the SAM is hydrolyzed over the time course of the crystallization. The conformation of SAH bound to all the class III methyltransferases is remarkably similar, with less than 0.230 Å root mean square deviation over all of the SAH residues (Fig. 4, *a–d*).

Tetrapyrrole Binding Site—The proposed binding site for the substrate cobalt factor II is a trough in the N-terminal domain, surrounded by several flexible loops. The trough is framed by loops between $\beta 2$ and αC and between $\beta 3$ and αD in the N-terminal domain and between $\beta 6$ and αG , between $\beta 7$ and αH , and the β -turn between $\beta 9$ and $\beta 10$ in the C-terminal domain. The bottom of the trough is formed by the N-terminal loop between $\beta 4$ and αE . Seven conserved residues can be found within these loops: Pro³⁵ ($\beta 2$ - αC), Ser¹⁰⁹ and Thr¹¹⁰ ($\beta 4$ - αE), Val¹⁵⁵ ($\beta 6$ - αG), Lys¹⁷⁶ ($\beta 7$ - αH), Tyr²²⁰, and Ala²²² ($\beta 9$ - $\beta 10$) (Fig. 5). These residues are specific to CbiL and CobI and are not found in the other class III methyltransferases, apart from A222, which is present in *B. megaterium* CbiF. Pro³⁵, Ser¹⁰⁹, and Thr¹¹⁰ are only conserved in CbiL.

To investigate the binding of the tetrapyrrole-derived substrate further, cobalt factor II was prepared in crystallization buffer as described under “Experimental Procedures.” CbiL crystals were soaked in substrate anaerobically for 10 min and then harvested. Longer incubation times were attempted, but this led to the destruction of the crystals. From these seven soaks, electron density maps were generated, of which three showed a similar ring of density with a central sphere of density in the putative substrate-binding site, suggesting a tetrapyrrole-like ring structure with a chelated metal ion. This tetrapyrrole-like density is in the A-subunit (Fig. 5*a*; crystal soak seven, 1.9 Å resolution). Attempts to refine the structure did not enhance the electron density, so this model must be seen as speculative, only poorly accounting for the multiple binding modes present in the crystal.

The electron density for cobalt factor II is weak, and there is no density for the acetate and propionate side chains of the substrate; further, the position of the doughnut of density places two of the side chains in steric conflict with the extended $\beta 2$ - $\alpha 3$ Mth CbiL loop in the major conformation it adopts in the crystal. There is additional density below the ring, suggesting that there may be a second deeper binding mode for the

the acetate and propionate side chains, and there is a steric conflict between the tetrapyrrole density and the $\beta 2$ - $\alpha 3$ loop. *b*, position of the tetrapyrrole within the *M. thermautotrophicus* CbiL architecture. Lys¹⁷⁶, Tyr²²⁰, and SAH are shown in a stick representation. *c*, close-up of the binding site; the distance from C20 to the sulfur of SAH is 5.5 Å, and the geometry is correct for methylation. The roles of Lys¹⁷⁶ and Tyr²²⁰ are discussed under “Results.”

Structure and Function of CbiL

tetrapyrrole. However, the doughnut of density with a blob at the center would place the tetrapyrrole close to the sulfur of SAH and in a position appropriate for methyl transfer to one of the carbon atoms. In the model, C20 was placed adjacent to the SAH with a sulfur to C20 distance of 5.5 Å, and it is a similar distance from C20 to the oxygen atom of the Tyr²²⁰ side chain. Tetrapyrrole binding is compatible with the following observations. (i) The $\beta 2$ - $\alpha 3$ loop is one of the more flexible parts of the *M. thermotrophicus* CbiL structure and has a different conformation in the A and B subunits due to the different environments experienced in the crystal. (ii) Attempts to increase the occupancy of the tetrapyrrole by increasing its concentration lead to crystal damage and loss of diffraction, a result compatible with movement of a part of the *M. thermotrophicus* CbiL structure in response to binding. (iii) No additional density was seen in this region for the nonsoaked crystals. Cocrystallization would offer a means of trapping a complex at higher occupancy, but to date such trials, necessarily under anaerobic conditions, have failed to yield crystals.

Analysis of SAH and Cobalt Factor II Binding Using EPR Spectroscopy—The EPR spectrum of Co(II) factor II shows a typical low spin, square planar cobalt environment exhibiting an axial spectrum with g_{\perp} (perpendicular) of 2.52 and cobalt hyperfine splitting of 114G (Fig. 6a) (30). Upon the addition of excess (10 eq) SAH, g_{\perp} shifts to 2.28, indicating the addition of a fifth ligand to the Co(II) ion (Fig. 6b) (30). Since the only difference between these samples is the addition of SAH, the source of the fifth ligand must be the SAH. There is no obvious superhyperfine splitting of the g_{\parallel} features of the spectrum, which would indicate that the ligating atom was a nitrogen, but the absence of such splitting does not exclude nitrogen as a ligand. Hence the carboxyl group, nitrogen atoms (from either the homocysteine or adenine moieties), or possibly the sulfur atom of SAH are all potential ligating atoms. The spectrum of the stoichiometric mixture of Co(II) factor II and wild type CbiL (Fig. 6c) is very similar, but not identical, to that of “free” Co(II) factor II. The differences are induced by binding to CbiL and are most evident around 2700 G. The addition of excess SAH with the intention of forming the ternary CbiL-Co(II) factor II-SAH gives rise to an EPR spectrum indicative of five coordinate Co(II) with $g_{\perp} = 2.29$, (Fig. 6d). This spectrum differs, however, from that observed in the binary Co(II) factor II-SAH complex (Fig. 6b), especially in the region between 2700 and 3100 G. These differences suggest that the spectrum in Fig. 6d does arise from the ternary complex and that the fifth ligand to the cobalt II ion may not be SAH in the ternary complex. The protein may supply an alternative fifth ligand to the cobalt ion, the ligand only moving into place on formation of the ternary complex and not as CbiL binds Co(II) factor II. Such “gating” may inhibit methyl group transfer to water in the absence of factor II. The observation that the x-ray structure does not show a protein ligand to the bound cobalt factor II may be explained by the method of production, the EPR samples being made in solution while the x-ray structure was produced by soaking preexisting crystals.

Investigating the Role of D104A, K176A, and Y220A in the Activity of CbiL—In order to probe the *in vivo* activities of the CbiL variants D104A, K176A, and Y220A, a similar functional

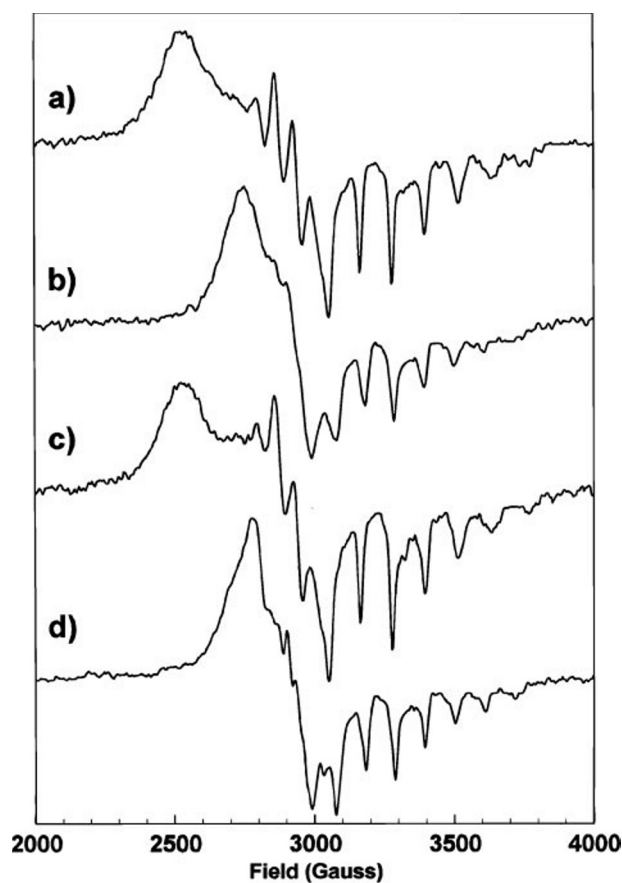


FIGURE 6. EPR spectra of Co(II) factor II in solution and bound to wild type CbiL. a, Co(II) factor II. b, Co(II) factor II plus 13 equivalents of SAH. c, Co(II) factor II plus 1 eq of wild type CbiL. d, Co(II) factor II plus 1 eq of wild type CbiL and 13 eq of SAH. All solutions were in 50 mM HEPES buffer, pH 8.0. Experimental conditions are given under “Experimental Procedures.”

complementation study as used for activity determination of wild type *M. thermotrophicus* CbiL was carried out through the complementation of the *S. enterica cbiL* strain. None of these variants were found to be capable of restoring cobalamin biosynthesis, indicating that all three were catalytically inactive *in vivo*.

To study these mutant proteins *in vitro*, the three CbiL variants were overproduced as recombinant proteins using the methods previously described for wild type protein. During purification, the K176A and Y220A variants behaved in a similar way to wild type CbiL, whereas the D104A mutant protein has a tendency to aggregate, suggesting that the mutation makes the protein less stable. The activities of the isolated protein variants were measured. Only the D104A mutant showed *in vitro* activity, although its rate was significantly less than wild type CbiL. Crystals of the K176A mutant, isomorphous with those of wild type CbiL, were grown but diffracted significantly less well than wild type CbiL. The σ A-weighted Fourier synthesis at 2.5 Å showed that the lysine side chain was missing, but there were no other structural changes; therefore, the loss of activity observed for the K176A mutant is a direct effect of the removal of the side chain of Lys¹⁷⁶.

Using ITC, K_D values were determined for SAM and SAH binding (Table 3). The most remarkable observation is the much higher affinity of SAH for the enzyme. As anticipated, the

D104A mutant has lower affinity for SAM. In fact, it was difficult to measure using ITC. Despite this, the D104A variant still bound SAH with high avidity, with a K_D similar to that of wild type enzyme (Fig. 7).

Analysis of SAH and Cobalt Factor II Binding to Mutant CbiL Using EPR Spectroscopy—Spectra of the binary complexes formed using both wild type (Fig. S2a) and mutant CbiL (Fig. S2, b–d) and Co(II) factor II were recorded by EPR. All of these spectra are very similar to one another, with $g_{\perp} = 2.52$, and show slight differences when compared with Fig. 6a. This suggests that Co(II) factor II binds to the mutant CbiL proteins in the same manner as it binds to the wild type and that the cobalt ion remains four-coordinate upon binding. The addition of excess SAH gives rise to the EPR spectra shown in Fig. 8. The spectra of the samples containing mutant CbiL (Fig. 8, b–d), are very similar to those observed in Fig. 6b, which arises from the binary Co(II) factor II-SAH complex with $g_{\perp} = 2.28$. They do not show the changes observed in Fig. 6d, which we interpret as

being indicative of ternary complex formation with CbiL. This suggests that SAH cannot bind to the mutant proteins in the presence of Co(II) factor II but rather binds to the free Co(II) factor II, providing a fifth ligand to the cobalt ion. This five-coordinate species will not bind to the CbiL mutants at the concentrations employed in the experiment; hence, the EPR spectrum of the “free” Co(II) factor II-SAH complex is observed.

DISCUSSION

In an attempt to understand why nature transiently methylates the C20 position of the tetrapyrrole framework during corrin biosynthesis, Eschenmoser (8) demonstrated that a secondary alcohol at this position, such as that found in precorrin-3B, had little chance of promoting ring contraction, whereas a tertiary alcohol proceeded in good yield. In essence, methylation at C20 is essential to prevent irreversible tautomerization to an oxo-derivative. Nature therefore employs an enzyme (CobI or CbiL) to ensure that this position is specifically methylated, although this methylated carbon position is subsequently lost during the biosynthesis. To determine more about the function and specificity of this enzyme, structural and mechanistic investigations on CbiL were undertaken.

The *cbiL* gene from *M. thermoautotrophicus* was cloned, and the function of the encoded protein was confirmed as a C20 methyltransferase of cobalamin biosynthesis by its ability to complement a defined *S. enterica cbiL* strain and the activity of the purified protein to methylate Co(II) factor II. Moreover, the preference of CbiL for cobalt factor II over other potential substrates, such as cobalt-precorrin-2

and factor II or precorrin-2, provides further evidence that cobalt factor II is a true pathway intermediate for cobalamin biosynthesis. It also indicates that cobalt factor III is an intermediate on the pathway, which is significant, since there is some discrepancy as to the identity of the oxidation state of the substrate for the next pathway enzyme (11, 31).

The oxidation state of the cobalt ion does not change when Co(II) factor II becomes methylated anaerobically by *M. thermoautotrophicus* CbiL. Therefore, CbiL does not alter the oxidation state of the cobalt ion, suggesting that cobalt does not play a redox role in the ring contraction process (25), which follows from the synthesis of cobalt factor III (2, 11, 31, 32).

The crystal structure of the enzyme was solved, and a comparison with other type III methyltransferases reveals that whereas the central β -sheets of the two domains are

TABLE 3

	K_D
	M
SAM binding	
Wild type CbiL	$6.71 \times 10^{-4} \pm 8.0 \times 10^{-6}$
D104A	Not available
K176A	$8.55 \times 10^{-5} \pm 1.73 \times 10^{-6}$
Y220A	$5.21 \times 10^{-4} \pm 7.12 \times 10^{-6}$
SAH binding	
Wild type CbiL	$6.45 \times 10^{-7} \pm 2.5 \times 10^{-8}$
D104A	$2.78 \times 10^{-6} \pm 5.69 \times 10^{-7}$
K176A	$1.50 \times 10^{-6} \pm 6.7 \times 10^{-8}$
Y220A	$1.30 \times 10^{-6} \pm 1.02 \times 10^{-7}$

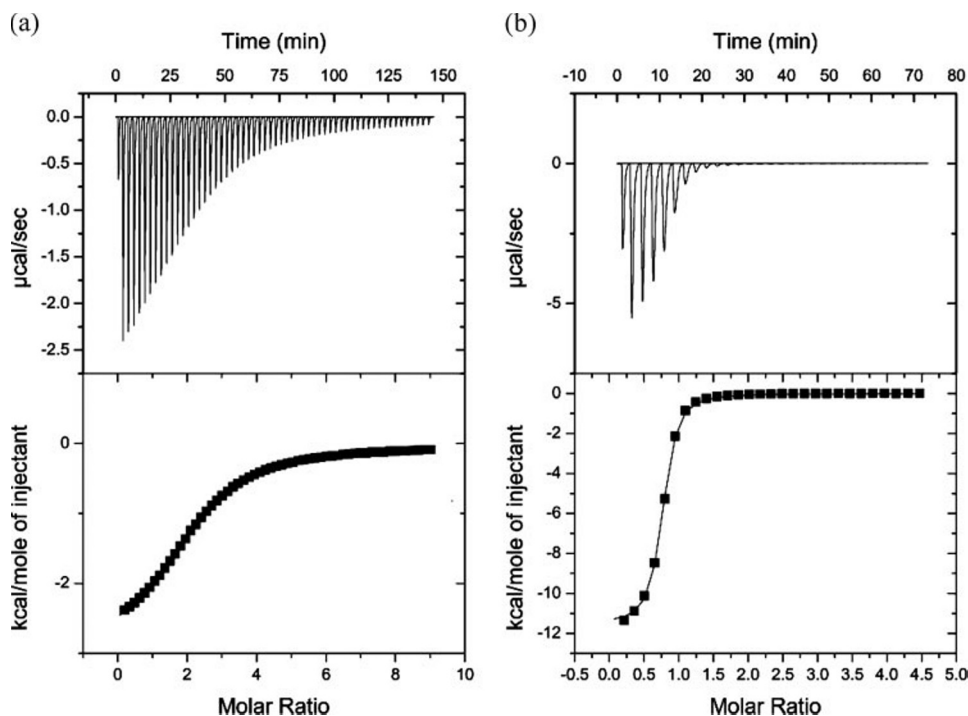


FIGURE 7. *a*, ITC profile for SAM binding to CbiL variant K176A. *Top*, base line-subtracted raw titration data obtained for injections of SAM into the protein solution; *bottom*, the integrated areas for the peaks normalized to the cofactor/protein molar ratio. *Dark dots* showing the experimental data and *solid lines* showing the best fit to one set of binding sites model were obtained by nonlinear least squares fitting. *b*, ITC profile for SAH binding to CbiL mutant K176A.

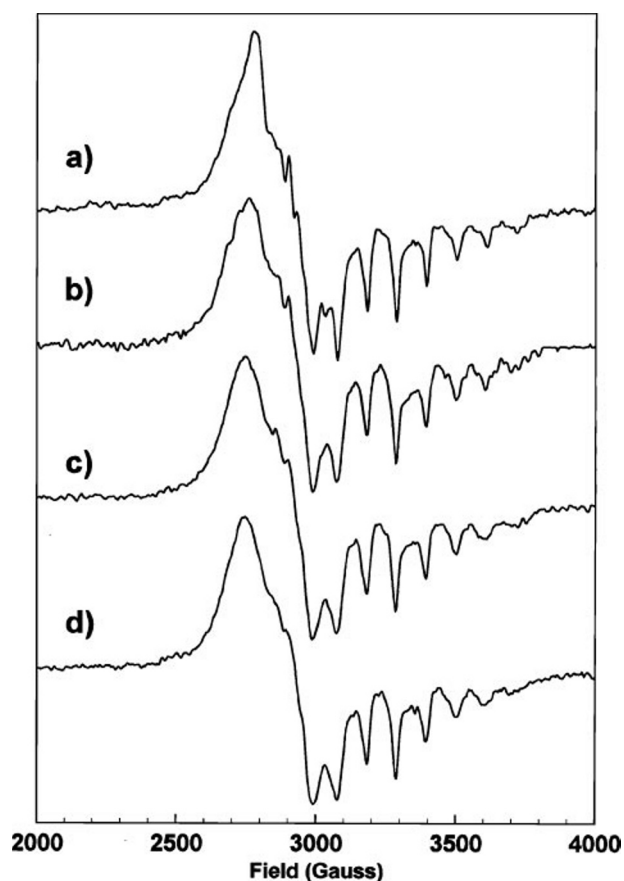


FIGURE 8. EPR spectra of Co(II) factor II, SAH, and mutant CbiL. All solutions contain 1 eq of Co(II) factor II, 13 eq of SAH, and 1 eq of CbiL. *a*, wild type CbiL for comparison (see Fig. 6*d*). *b*, D104A mutant. *c*, K176A mutant. *d*, Y220A mutant.

relatively well conserved, the number and position of the surrounding helices are highly variable (27). Some of the variable loops are likely to contribute to substrate specificity (Fig. 4). There are also significant differences from the recently published structure of the putative CbiL from *C. tepidum*, most significantly in the N-terminal domain, where $\alpha 3$ is considerably extended in the *M. thermoautotrophicus* CbiL, and in the C-terminal domain, where there are different numbers of helices above and below the β -sheet plane (12).

The crystallized enzyme has one SAH bound per subunit. The crystals were also soaked with the enzyme's tetrapyrrole-derived substrate, cobalt factor II. From these soaking experiments, evidence of a substrate complex was obtained, whereby weak electron density was observed within the active site. The proximity of the density to the sulfur of the SAH suggests that the complex may be real and represents the first complex observed for any of the class III enzymes. This allowed modeling of cobalt factor III into the active site in an orientation that would permit catalysis, such that the C20 position was placed 5.5 Å away from the sulfur of SAH. This also readily identified a number of important amino acid residues that are likely to be involved in either catalysis or substrate recognition. In support of their role, these amino acid residues, Tyr²²⁰ and Lys¹⁷⁶, are found to be invariant among the C20 methyltransferases. Sequence analysis had previously identified three conserved residues that were likely to play an important role in the func-

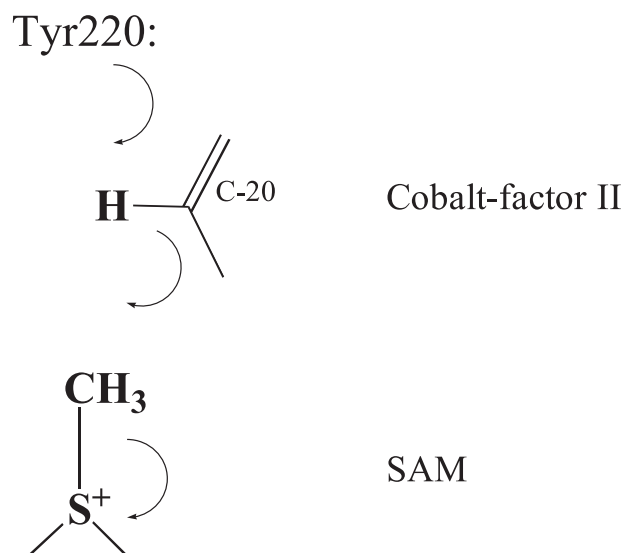


FIGURE 9. Potential catalytic role for Tyr²²⁰. The side chain (base) might facilitate a quasi-S_N2-like mechanism of methyl group transfer from SAM to C20 of cobalt factor II.

tion of the enzyme. A further amino acid, Asp¹⁰⁴, was found to be invariant among all of the class III enzymes. The roles played by these amino acids were investigated by site-directed mutagenesis.

The mutant proteins were tested for *in vivo* and *in vitro* activity and SAM and SAH binding ability. Due to its proximity to the SAM binding site, it was predicted that any modification of Asp¹⁰⁴ would affect SAM binding. Surprisingly, although D104A lost its ability to bind SAM, the enzyme demonstrated some *in vitro* activity. The disparity between the *in vivo* and *in vitro* data is likely to be explained by the high level of SAM used in the assay, allowing some productive binding with the enzyme variant, whereas the much lower physiological SAM concentrations prevented any *in vivo* activity from being observed. It is apparent that the negative charge associated with the residue is important for SAM binding, possibly in helping to attract the positive charge associated with the sulfonium ion of the SAM.

Lys¹⁷⁶ is in a position to help assist in the orientation of macrocyclic substrate binding, ensuring that it binds correctly by modulating specificity through interaction with the propionate side chains found on ring A or possibly D if the tetrapyrrole was flipped. The K176A crystal structure also revealed that there is no conformational change due to the loss of the side chain by the mutation. This strongly suggests that Lys¹⁷⁶ is indeed involved in binding cobalt factor II in the correct position for productive methyl transfer to take place. Tyr²²⁰ is likely to be involved in catalysis of the methyl group transfer from SAM to C20 of cobalt factor II. This could be facilitated by the side chain, acting as a base in an extended quasi-S_N2-like (bimolecular nucleophilic substitution) mechanism (Fig. 9). Indeed, in the substrate complex structure, this residue is found to be close enough to C20 of cobalt factor II to have such a theoretical function.

EPR analysis suggests that cobalt factor II can bind to free CbiL but that the cobalt remains four-coordinate. Upon the addition of SAH, the cobalt becomes five-coordinate, indicat-

ing that the protein is likely to supply a fifth ligand. In the CbiL structure from *C. tepidum*, it is proposed that Thr¹¹² may supply a ligand to the cobalt-containing substrate (12). However, in the *M. thermoautotrophicus* CbiL-cobalt factor II complex, the equivalent residue, Thr¹¹⁰, is too far away to act as a ligand. Nonetheless, if the substrate were to bind lower in the pocket, and there is some evidence in the electron density maps for deeper binding modes, then Thr¹¹⁰ would provide the fifth ligand to the cobalt ion.

In summary, we have conducted the first detailed structure/function analysis of the C20 methyltransferase of cobalamin biosynthesis. We have shown that the true substrate for the enzyme is Co(II) factor II and that the product of the reaction is Co(II) factor III. From our studies, we can infer that the oxidation state of the cobalt ion is not important for ring contraction, that substrate discrimination is imparted by a cobalt ligand in the enzyme, and that productive binding of a ternary complex prevents methyl transfer to water.

REFERENCES

- Blanche, F., Cameron, B., Crouzet, J., Debussche, L., Thibaut, D., Vuilhorgne, M., Leeper, F. J., and Battersby, A. R. (1995) *Angew. Chem. Int. Ed. Engl.* **34**, 383–411
- Roessner, C. A., and Scott, A. I. (2006) *J. Bacteriol.* **188**, 7331–7334
- Roth, J. R., Lawrence, J. G., and Bobik, T. A. (1996) *Annu. Rev. Microbiol.* **50**, 137–181
- Warren, M. J., Raux, E., Schubert, H. L., and Escalante-Semerena, J. C. (2002) *Nat. Prod. Rep.* **19**, 390–412
- Blanche, F., Thibaut, D., Debussche, L., Hertle, R., Zipfel, F., and Müller, G. (1993) *Angew. Chem. Int. Ed. Engl.* **32**, 1651–1653
- Roth, J. R., Lawrence, J. G., Rubenfield, M., Kieffer-Higgins, S., and Church, G. M. (1993) *J. Bacteriol.* **175**, 3303–3316
- Raux, E., Schubert, H. L., Roper, J. M., Wilson, S. W., and Warren, M. J. (1999) *Bioorganic Chem.* **27**, 100–118
- Eschenmoser, A. (1988) *Angew. Chem. Int. Ed. Engl.* **27**, 5–39
- Roessner, C. A., Spencer, J. B., Ozaki, S., Min, C., Atshaves, B. P., Nayar, P., Anousis, N., Stolowich, N. J., Holderman, M. T., and Scott, A. I. (1995) *Protein Expression Purif.* **6**, 155–163
- Spencer, P., Stolowich, N. J., Sumner, L. W., and Scott, A. I. (1998) *Biochemistry* **37**, 14917–14927
- Santander, P. J., Kajiwarra, Y., Williams, H. J., and Scott, A. I. (2006) *Bioorg. Med. Chem.* **14**, 724–731
- Wada, K., Harada, J., Yaeda, Y., Tamiaki, H., Oh-Oka, H., and Fukuyama, K. (2007) *FEBS J.* **274**, 563–573
- Raux, E., Leech, H. K., Beck, R., Schubert, H. L., Santander, P. J., Roessner, C. A., Scott, A. I., Martens, J. H., Jahn, D., Thermes, C., Rambach, A., and Warren, M. J. (2003) *Biochem. J.* **370**, 505–516
- Raux, E., Lanois, A., Levillayer, F., Warren, M. J., Brody, E., Rambach, A., and Thermes, C. (1996) *J. Bacteriol.* **178**, 753–767
- Leslie, A. G. W. (1992) *Joint CCP4 + ESF-EAMCB Newsletter Protein Crystallogr.* **26**
- Collaborative Computational Project, N. (1994) *Acta Crystallogr. Sect. D* **50**, 760–763
- Vagin, A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30**, 1022–1025
- Otwinowski, Z. (1991) in *Isomorphous Replacement and Anomalous Scattering* (Wolf, W., and Leslie, A.G.W. eds) pp. 80–86, Daresbury, CCP4 Study Weekend Proceedings, SERC, Daresbury Laboratory, Warrington, U.K.
- Cowtan, K. D., and Zhang, K. Y. (1999) *Prog. Biophys. Mol. Biol.* **72**, 245–270
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255
- Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47**, 110–119
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
- Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) *Anal. Biochem.* **179**, 131–137
- Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J. N., et al. (1997) *J. Bacteriol.* **179**, 7135–7155
- Scott, A. I. (1994) *Heterocycles* **39**, 471–476
- Schubert, H. L., Wilson, K. S., Raux, E., Woodcock, S. C., and Warren, M. J. (1998) *Nat. Struct. Biol.* **5**, 585–592
- Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003) *Trends Biochem. Sci.* **28**, 329–335
- Stroupe, M. E., Leech, H. K., Daniels, D. S., Warren, M. J., and Getzoff, E. D. (2003) *Nat. Struct. Biol.* **10**, 1064–1073
- Vevodova, J., Graham, R. M., Raux, E., Schubert, H. L., Roper, D. I., Brindley, A. A., Ian Scott, A., Roessner, C. A., Stamford, N. P., Elizabeth Stroupe, M., Getzoff, E. D., Warren, M. J., and Wilson, K. S. (2004) *J. Mol. Biol.* **344**, 419–433
- Bayston, J. H., Looney, F. D., Pilbrow, J. R., and Winfield, M. E. (1970) *Biochemistry* **9**, 2164–2172
- Kajiwarra, Y., Santander, P. J., Roessner, C. A., Perez, L. M., and Scott, A. I. (2006) *J. Am. Chem. Soc.* **128**, 9971–9978
- Santander, P. J., Roessner, C. A., Stolowich, N. J., Holderman, M. T., and Scott, A. I. (1997) *Chem. Biol.* **4**, 659–666
- Brindley, A. A., Raux, E., Leech, H. K., Schubert, H. L., and Warren, M. J. (2003) *J. Biol. Chem.* **278**, 22388–22395
- Raux, E., McVeigh, T., Peters, S. E., Leustek, T., and Warren, M. J. (1999) *Biochem. J.* **338**, 701–708