

# Structure and possible mechanism of the CcbJ methyltransferase from *Streptomyces caelestis*

Jacob Bauer,<sup>a\*</sup> Gabriela Ondrovičová,<sup>a</sup> Lucie Najmanová,<sup>b</sup> Vladimír Pevala,<sup>a</sup> Zdeněk Kameník,<sup>b</sup> Július Košťan,<sup>c</sup> Jiří Janata<sup>b</sup> and Eva Kutejová<sup>a,b\*</sup>

<sup>a</sup>Institute of Molecular Biology, Slovak Academy of Sciences, 851 45 Bratislava, Slovakia,

<sup>b</sup>Institute of Microbiology, Academy of Sciences of the Czech Republic, 142 20 Prague, Czech Republic, and <sup>c</sup>Department for Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, A-1030 Vienna, Austria

Correspondence e-mail: jacob.bauer@savba.sk, evakutej@hotmail.com

The *S*-adenosyl-L-methionine (SAM)-dependent methyltransferase CcbJ from *Streptomyces caelestis* catalyzes one of the final steps in the biosynthesis of the antibiotic celesticetin, methylation of the N atom of its proline moiety, which greatly enhances the activity of the antibiotic. Since several celesticetin variants exist, this enzyme may be able to act on a variety of substrates. The structures of CcbJ determined by MAD phasing at 3.0 Å resolution, its native form at 2.7 Å resolution and its complex with *S*-adenosyl-L-homocysteine (SAH) at 2.9 Å resolution are reported here. Based on these structures, three point mutants, Y9F, Y17F and F117G, were prepared in order to study its behaviour as well as docking simulations of both CcbJ–SAM–substrate and CcbJ–SAH–product complexes. The structures show that CcbJ is a class I SAM-dependent methyltransferase with a wide active site, thereby suggesting that it may accommodate a number of different substrates. The mutation results show that the Y9F and F117G mutants are almost non-functional, while the Y17F mutant has almost half of the wild-type activity. In combination with the docking studies, these results suggest that Tyr9 and Phe117 are likely to help to position the substrate for the methyl-transfer reaction and that Tyr9 may also facilitate the reaction by removing an H<sup>+</sup> ion. Tyr17, on the other hand, seems to operate by helping to stabilize the SAM cofactor.

Received 13 November 2013

Accepted 16 December 2013

**PDB references:** CcbJ, SeMet, 4hgy; native, 4hgz; SAH complex, 4hh4

## 1. Introduction

The constant need for new, biologically active lead compounds fuels an intensive ongoing search for new chemicals as well as efforts to produce directed modifications of presently known chemicals. The production of modified biologically active compounds frequently involves the genetic modification of appropriate biosynthetic clusters, the construction of hybrid biosynthetic clusters, precursor-directed biosynthesis and mutasynthesis (Weissman, 2007). Whereas the first two approaches result in the biosynthesis of modified precursors, in the latter two applications modified precursor molecules are supplied to a pre-existing biosynthetic pathway. For all of these applications, an important consideration is the substrate specificities of both the enzyme catalyzing the incorporation of the modified precursor and every enzyme which follows it in the pathway. In order to select the best target compounds and to design the best mutants of the most appropriate enzymes, it is necessary to have in hand a detailed biochemical characterization of these enzymes. Knowledge of the three-dimensional structure gives the most detailed characterization of an enzyme, and is important for understanding its mechanism of action as well as designing potential target modifications.

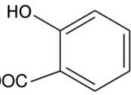
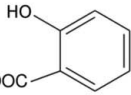
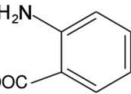
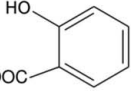
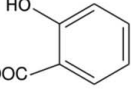
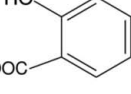
The lincosamide antibiotics are a small but clinically important group of antibacterial compounds with only two natural representatives: lincomycin and celesticetin (for a review, see Spížek & Řezanka, 2004b). Synthetic lincomycin derivatives exhibiting higher antibacterial activities (some of which also exhibited antiplasmodial activities) have been described (Magerlein, 1977); however, only one of them, the semisynthetic derivative clindamycin, is in clinical use. Clindamycin is more potent than lincomycin and is the most widely used of the lincosamides. The mutasynthetic preparation of more active lincomycin derivatives has been described previously (Ulanova *et al.*, 2010), and clindamycin derivatives with antiprotozoal activities can potentially also be prepared.

The lincosamide core structure consists of two building blocks: an amino sugar and an amino acid, either proteinogenic L-proline (hereafter referred to as proline) in celesticetin or 4-propyl-L-proline (hereafter referred to as propylproline) in lincomycin. These two units are synthesized separately before being joined together through an amide bond. Natural modifications to this core structure include the attachment of a salicylate unit (in celesticetin only) and a number of methylations. The only methylation common to both celesticetin and lincomycin is the N-methylation of the amino-acid unit, which occurs after condensation of the two units. The N-methylation of *N*-demethylincomycin is the last step in lincomycin

biosynthesis (Chung *et al.*, 1997) and is catalyzed by the protein LmbJ (Najmanová *et al.*, 2013). An analogous reaction should be catalyzed in celesticetin biosynthesis by the LmbJ orthologue CcbJ. This methylation has been shown to enhance the activity of lincomycin 4–15-fold (Magerlein, 1977).

Both LmbJ and CcbJ have been biochemically characterized (Najmanová *et al.*, 2013). Although *N*-demethylincomycin has an *n*-propyl group attached to the amino-acid moiety which is missing in the corresponding celesticetin precursor, both LmbJ and CcbJ were able to catalyze the *N*-methylation of *N*-demethylincomycin *in vitro* with similar kinetic parameters. This indicates that CcbJ does not have a very tight substrate specificity. Previous mutasynthesis experiments had established that LmbJ also has a weaker substrate specificity (Ulanova *et al.*, 2010). This loose substrate specificity could greatly facilitate the adaptation of both biosynthetic pathways to the production of newly designed, clinically important compounds containing modified amino-acid moieties.

The biochemical results established that CcbJ is a SAM-dependent methyltransferase. The present work shows unambiguously that it belongs to the class I family of SAM-dependent methyltransferases. Methylation is a common modification reaction in the biosynthesis of specialized metabolites and can significantly alter the biological function

Name	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$
Lincomycin	SCH <sub>3</sub>	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	OH	H
<i>N</i> -Demethylincomycin	SCH <sub>3</sub>	H	CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	OH	H
Celesticetin		CH <sub>3</sub>	H	OCH <sub>3</sub>	H
Celesticetin B	SCH <sub>2</sub> CH <sub>2</sub> OOC- 	CH <sub>3</sub>	H	OCH <sub>3</sub>	H
Celesticetin C	SCH <sub>2</sub> CH <sub>2</sub> OOCCH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	CH <sub>3</sub>	H	OCH <sub>3</sub>	H
Celesticetin D		CH <sub>3</sub>	H	OCH <sub>3</sub>	H
Desalicyletin	SCH <sub>2</sub> CH <sub>2</sub> OOCCH <sub>3</sub>	CH <sub>3</sub>	H	OCH <sub>3</sub>	H
<i>N</i> -Demethylcelesticetin	SCH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>3</sub>	H	OCH <sub>3</sub>	H
<i>N,O</i> -Didemethylcelesticetin		H	H	H	H
Desalicyletinsalicylate		CH <sub>3</sub>	H	OCH <sub>3</sub>	H
Clindamycin	SCH <sub>2</sub> CH <sub>2</sub> OOC-  SCH <sub>3</sub>	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	OH	Cl

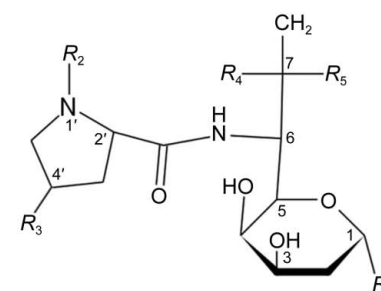


Figure 1

Selected lincosamide antibiotics. The core lincosamide structure is shown on the right; the table details various modifications. CcbJ (or LmbJ) adds a methyl group to the  $R_2$  substituent. It can be seen that lincomycin adds an *n*-propyl group to the proline moiety at the  $R_3$  substituent, replaces the methyl ether at  $R_4$  with a hydroxyl group and replaces the salicylic thioether at  $R_1$  with a methyl thioether. The various forms of celesticetin which have been isolated from *S. caelestis* fermentation media include those which replace the salicylate group with an aminosalicylate group (celesticetin C), a *para*-aminosalicylate (desalicyletinsalicylate) group, an alcohol (desalicyletin), an acetate (celesticetin D) and an isobutyl ester (celesticetin B) (reviewed in Spížek & Řezanka, 2004a).

**Table 1**  
Crystal properties.

	SeMet CcbJ	Native CcbJ	CcbJ-SAH
Space group	C222 <sub>1</sub>	C222 <sub>1</sub>	C222 <sub>1</sub>
Unit-cell parameters (Å)	<i>a</i> = 169.95, <i>b</i> = 244.00, <i>c</i> = 118.07	<i>a</i> = 168.02, <i>b</i> = 244.55, <i>c</i> = 117.85	<i>a</i> = 169.95, <i>b</i> = 244.00, <i>c</i> = 118.07
No. of monomers in asymmetric unit	6	6	6
<i>V<sub>M</sub></i> (Å <sup>3</sup> Da <sup>-1</sup> )	3.63	3.63	3.63
Solvent content (%)	66	66	66

of the final product (for a review, see Struck *et al.* 2012). Methyltransferases are therefore potentially useful in biotechnological applications and for work in emerging areas of chemical biology and synthetic biology such as metabolic engineering. Structurally, the class I methyltransferases are characterized by a common core fold consisting of an open, twisted mixed seven-stranded  $\beta$ -sheet flanked by  $\alpha$ -helices; strand 7 of this sheet is antiparallel to the other six strands. This core fold is frequently modified in a number of ways depending on the particular type of substrate being modified. For example, small-molecule methyltransferases normally have an active-site cover which sits over the substrate-binding site, while those which modify larger substrates often have additions to their C-terminal ends (for a recent review, see Liscombe *et al.*, 2012). Despite this structural conservation, the sequence identity in the methyltransferase core fold is often only 15–20% for small-molecule methyltransferases. The only conserved areas common to all class I methyltransferases are a GxGxG motif found at the end of  $\beta$ 1 and an acidic residue, either glutamate or aspartate, at the end of  $\beta$ 2. One consequence of this lack of sequence conservation is that the SAM cofactor is bound in somewhat different ways by different methyltransferases. Common to all of them is that the conserved acidic residue is always hydrogen-bonded to the O2' and O3' hydroxyl groups of the cofactor, while the backbone amide N atoms of the GxGxG motif frequently, but not invariably, form hydrogen bonds to the carboxy group of the cofactor (Fauman *et al.*, 1999).

The mechanism of catechol *O*-methyltransferase, the simplest of the class I methyltransferases, has been studied in detail by Hegazi *et al.* (1976, 1979) and Woodard *et al.* (1980). The results of both groups support the idea that the methyl group is transferred from the cofactor to the substrate *via* an S<sub>N</sub>2 mechanism. This mechanism is generally thought to apply to all class I methyltransferases with nucleophilic N, O and S target atoms.

Although it has been unambiguously established that LmbJ catalyzes the final step in lincomycin biosynthesis (Chung *et al.*, 1997), it is not at all clear that N-methylation of the proline moiety is the final step in celesticetin biosynthesis. It may precede the *O*-methylation of the amino sugar or condensation with the salicylate moiety. Alternatively, each of these reactions could proceed independently of the others. If N-methylation occurs after salicylate condensation, then this enzyme would be particularly interesting because it would act not just on its native substrate but also on the multitude of

naturally occurring celesticetin derivatives, some of which are shown in Fig. 1. To study the details of its reaction mechanism and specificity, we report here the three-dimensional structure of CcbJ determined by MAD phasing to 3.0 Å resolution and its native form at 2.7 Å resolution. We were also able to determine the structure of a CcbJ-SAH complex at 2.9 Å resolution. On the basis of these structures, we also prepared three point mutants to study its activity as well as docking simulations of both substrate and product complexes. These results provide a sound basis for the further understanding of this enzyme and its various characteristics.

## 2. Materials and methods

### 2.1. Expression-construct preparation

The expression system for the production of recombinant N-terminally His-tagged CcbJ (UniProt ID E9JES0) using the pET-28b expression vector has been described in a previous paper (Najmanová *et al.*, 2013).

The CcbJ point mutants were produced using QuikChange XL site-directed mutagenesis (Stratagene) following the instructions of the manufacturer. The oligonucleotide primers listed in Supplementary Table S1<sup>1</sup> were used to introduce a single point mutation that was verified by DNA sequencing. The resulting CcbJ mutants were fusion proteins that carried the same amino-terminal hexahistidine tag and linker as described above.

### 2.2. Expression of recombinant CcbJ

Native CcbJ carrying an N-terminal His tag was overproduced in *Escherichia coli* BL21 (DE3) cells (Novagen). The bacterial cells were grown at 37°C in Luria–Bertani medium (1% Bacto tryptone, 0.5% yeast extract, 1% NaCl) with 50 µg ml<sup>-1</sup> kanamycin; protein overexpression was induced with 0.4 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) for 1 h. The cells were pelleted by centrifugation and were stored at –80°C. Selenomethionine-labelled CcbJ was produced in *E. coli* 834 (DE3) cells (Novagen) by the amino-acid saturation method (Doublé, 1997). Bacteria were grown at 37°C in minimal M9 medium supplemented with 0.4% glucose, 2 mM MgSO<sub>4</sub>, 25 µg ml<sup>-1</sup> FeSO<sub>4</sub>·7H<sub>2</sub>O, 1 µg ml<sup>-1</sup> vitamins (biotin and thiamine), 50 µg ml<sup>-1</sup> kanamycin, 40 µg ml<sup>-1</sup> each of the amino acids L-lysine, L-phenylalanine, L-threonine, L-isoleucine, L-leucine and L-valine (to inhibit L-methionine production) and L-selenomethionine at a concentration of 40 µg ml<sup>-1</sup>. Overproduction was induced with 0.4 mM IPTG for 3.5 h and the cells were stored at –80°C after centrifugation.

### 2.3. Purification of recombinant native or selenomethionyl CcbJ protein

Cells were resuspended in solubilization buffer [20 mM Tris–HCl pH 8.0, 150 mM NaCl, 20%(v/v) glycerol] and

<sup>1</sup> Supporting information has been deposited in the IUCr electronic archive (Reference: DW5086).

**Table 2**

Data-collection statistics.

Values in parentheses are for the highest resolution shell.

	SeMet CcbJ			Native CcbJ	CcbJ-SAH
	Peak	Inflection	Remote		
Wavelength (Å)	0.97936	0.97944	0.97549	0.97549	0.97549
Resolution range (Å)	70.0–3.00 (3.16–3.00)	70.0–3.00 (3.16–3.00)	70.0–3.00 (3.16–3.00)	54.59–2.70 (2.85–2.70)	73.36–2.90 (3.06–2.90)
Measured reflections	714343 (106013)	717576 (106464)	719014 (106582)	477177 (70435)	221500 (21477)
Unique reflections	49088 (7117)	49193 (7137)	49292 (7136)	66712 (9634)	52391 (6436)
Multiplicity	14.6 (14.9)	14.6 (14.9)	14.6 (14.9)	7.2 (7.3)	4.2 (3.3)
Completeness (%)	100 (100)	100 (100)	100 (100)	99.9 (100)	96.0 (81.6)
$\langle I/\sigma(I) \rangle$	26.7 (8.6)	26.0 (7.7)	24.9 (6.3)	17.6 (4.1)	15.8 (3.0)
$R_{\text{merge}}$ (%)	7.4 (29.7)	7.6 (36.9)	8.0 (47.0)	7.1 (47.2)	7.9 (41.3)

sonicated on ice. The cell lysate was centrifuged for 15 min at 100 000g and the supernatant was loaded onto an Ni<sup>2+</sup>-NTA affinity column (Qiagen). The column was washed with five column volumes of solubilization buffer containing 40 mM imidazole and the tagged protein was eluted in three steps with solubilization buffer containing 0.1, 0.2 or 0.3 M imidazole; the active fraction was eluted in the 0.2 M imidazole step. The buffer was then exchanged to 50 mM Tris-HCl pH 8.3 (50 mM Tris-HCl pH 8.3 with 1 mM dithiothreitol for the selenomethionyl derivative) using a NAP-10 column (Pharmacia). Fractions containing CcbJ were pooled, concentrated to 20 mg ml<sup>-1</sup> using a Centricon YM-10 filter (Millipore) and stored at -80°C in aliquots until needed. The protein concentration was estimated using the BCA protein assay kit (Pierce Biotechnology) and protein purity was verified using 15% SDS-PAGE and Coomassie Brilliant Blue staining.

#### 2.4. Protein crystallization

Crystals were grown using the hanging-drop vapour-diffusion method. The reservoir volumes were 0.5 ml and each drop was made by mixing 1.5 µl protein solution with 1.5 µl reservoir solution. Initial screening was carried out using Structure Screens 1 and 2 from Molecular Dimensions (Suffolk, England) and Crystal Screen and Crystal Screen 2 from Hampton Research (Aliso Viejo, California, USA). Promising crystals appeared in condition No. 27 of Structure Screen 1 (0.1 M Na HEPES pH 7.5, 1.5 M Li<sub>2</sub>SO<sub>4</sub>·H<sub>2</sub>O). Diffraction-quality crystals were obtained after optimization in 0.1 M Na HEPES pH 7.5, 1.3 M Li<sub>2</sub>SO<sub>4</sub>·H<sub>2</sub>O; somewhat larger crystals could be prepared if the crystallization solution also contained 10% glycerol or ethylene glycol. Native CcbJ crystals grew in 2–6 d at a temperature of 18°C and an initial protein concentration of 5–10 mg ml<sup>-1</sup>. Crystals of SeMet-labelled CcbJ grew in the same conditions in the same time at a protein concentration of 4–11 mg ml<sup>-1</sup>. Crystals of a CcbJ-S-adenosylhomocysteine complex (which was originally thought to contain CcbJ-S-adenosylmethionine) were prepared in the same conditions as the native protein with the addition of 4 mM SAM, but only appeared after 2–3 months at 18°C.

#### 2.5. Data collection and structure solution

Prior to data collection, all crystals were briefly soaked in a saturated solution of Li<sub>2</sub>SO<sub>4</sub> (2.7 M), which served as a cryoprotectant, and then flash-cooled in liquid nitrogen.

All data were collected at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) on beamline ID14-4 (McCarthy *et al.*, 2009) at 100 K. All crystals had approximate dimensions of 0.1 × 0.1 × 0.4 µm, belonged to space group C222<sub>1</sub> and had approximately the same Matthews coefficient ( $V_M$ ) of 3.63 Å<sup>3</sup> Da<sup>-1</sup>, which indicates that six monomers should be present in the asymmetric unit of each crystal. The unit-cell parameters for the SeMet CbbJ and the CcbJ-SAH crystals were identical ( $a = 169.95$ ,  $b = 244.00$ ,  $c = 118.07$  Å), while those of the native crystal were slightly different ( $a = 168.02$ ,  $b = 244.55$ ,  $c = 117.85$  Å). The crystal properties are summarized in Table 1.

A three-wavelength 3.00 Å resolution data set was collected from a SeMet-derivative crystal at wavelengths of 0.97936 Å (peak), 0.97944 Å (inflection point) and 0.97549 Å (high-energy remote). A 2.7 Å resolution single-wavelength data set was collected from a native crystal at a wavelength of 0.97549 Å, and a 2.9 Å resolution data set was collected from a CcbJ-SAH complex crystal at the same wavelength. All data were processed using *iMOSFLM* v.1.0.3 (Battye *et al.*, 2011), the space group was assigned using *POINTLESS* v.1.3.1 and the data were scaled with *SCALA* v.3.3.9 (Evans, 2006). Data-collection statistics for the three data sets can be seen in Table 2.

The structure of the SeMet derivative was solved by multiple anomalous dispersion phasing using the *CRANK* interface (Ness *et al.*, 2004) in *CCP4* v.6.1.13 (Winn *et al.*, 2011). *CRANK* ran the following programs: *TRUNCATE* (French & Wilson, 1978), *SCALEIT*, *AFRO* (for determining normalized substructure factor amplitudes), *CRUNCH2* (de Graaff *et al.*, 2001; substructure determination), *BP3* (Pannu *et al.*, 2003; Pannu & Read, 2004; MAD phasing) and *SOLOMON* (Abrahams & Leslie, 1996; density modification). 18 usable selenium positions were found and the overall figure of merit for the phasing was 0.484. The resulting maps were quite clear and could be fairly easily traced in most areas. 5% of the data was then set aside for  $R_{\text{free}}$  cross-validation and an initial model was constructed using 60 cycles of *Buccaneer*

**Table 3**  
Refinement statistics.

	SeMet CcbJ	Native CcbJ	CcbJ-SAH
Resolution range (Å)	70.0–3.00	54.6–2.70	73.4–2.90
Reflections			
In refinement	46589	63338	49744
In test set	2480	3353	2647
<i>R</i> factor	0.225	0.181	0.183
<i>R</i> <sub>free</sub>	0.248	0.224	0.221
No. of non-H atoms			
Total	10916	10930	12077
Protein	10753	10736	11663
Ligand (total)	72	59	310
SAH	0	0	156
Sulfate	55	41	101
Ethylene glycol	17	17	0
Glycerol	0	0	37
Li <sup>+</sup>	0	1	0
HEPES	0	0	16
Water molecules	92	136	107
Average <i>B</i> value (Å <sup>2</sup> )	62.2	63.2	43.7
Deviations from ideal			
Bond lengths (Å)	0.004	0.007	0.014
Bond angles (°)	0.747	1.03	1.50
Ramachandran plot (%)			
Favoured	96.8	96.8	97.0
Allowed	99.8	99.3	99.8
Outliers	0.22	0.72	0.20

(Cowtan, 2006, 2008) followed by *REFMAC5* v.5.5.0109 (Murshudov *et al.*, 1996, 1997, 2011; Pannu *et al.*, 1998; Skubák *et al.*, 2004). The completed initial model had an *R* factor of 0.260 and an *R*<sub>free</sub> of 0.311. After rebuilding in *O* v.12 (Jones *et al.*, 1991) and *Coot* (Emsley *et al.*, 2010), the model was again refined using *REFMAC5* with NCS restraints and TLS refinement (Winn *et al.*, 2001, 2003). The final model had an *R* factor of 0.225 and an *R*<sub>free</sub> of 0.248.

The native structure was solved by molecular replacement using *Phaser* v.2.1.4 (McCoy *et al.*, 2007) with the SeMet-derivative structure as a search model. After rebuilding and refinement, the *R* factor and *R*<sub>free</sub> were 0.181 and 0.224, respectively. The CcbJ-SAH complex structure was solved, rebuilt and refined in the same way. Its final *R* factor and *R*<sub>free</sub> were 0.183 and 0.221, respectively. It was at this point that it became apparent that the crystal contained not SAM but SAH. The *R*<sub>free</sub> test sets used were consistent across all data sets. All refinement statistics are summarized in Table 3.

## 2.6. Computational docking

Docking studies were carried out using *AutoDock Vina* v.1.1.2 (Trott & Olson, 2010). The CcbJ-SAH complex structure, a CcbJ-SAM complex derived from it and two mutant complex structures of CcbJ-SAM, Y9F and F117G, were used for the docking studies. The ligands used are listed in Supplementary Table S2. Five different base substrates were used, and four different forms of each base substrate were evaluated, making a total of 20 different ligands which were examined. The source for the lincomycin structures was the crystal structure of lincomycin reported by Rajeswaran & Srikrishnan (2004) and available from the Cambridge Crystallographic Data Centre (accession No. 235119). All celes-

**Table 4**  
Activities of wild-type CcbJ and mutants.

The uncertainty expresses the 95% confidence interval.

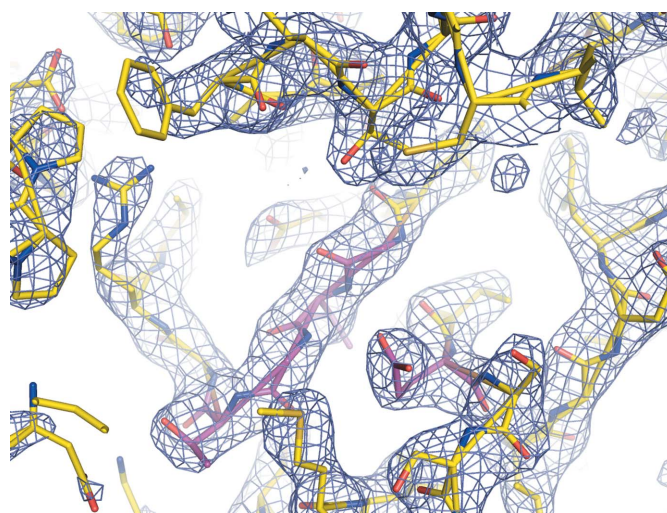
Mutant	Lincomycin (ng) per microgram of CcbJ
Native	199.6 ± 8.58
F117G	1.10 ± 1.76
Y9F	4.45 ± 2.73
Y17F	79.6 ± 2.15

ticetin structures were constructed by modifying this base structure using *PyMOL* v.1.3rc2 (Schrödinger). The final coordinates were obtained by submitting these structures to the *PRODRG* server (Schüttelkopf & van Aalten, 2004).

The ligands and complexes were prepared for docking by adding charges and H atoms and designating flexible torsion angles using *AutoDock Tools* v.1.5.6rc2 (Morris *et al.*, 2009). The same program was also used to determine the size and the location of the search box. Each docking search was performed with an exhaustiveness of 16, 30 modes and energy ranges of 5 kcal mol<sup>-1</sup>. For each ligand, four searches were carried out. Firstly, the docking was carried out over the whole molecule to confirm that the predicted binding site was in fact the correct one. Next, the search was carried out over only the binding site itself with all torsion angles in the ligand fixed. Thirdly, the search was conducted allowing the ligand to be flexible. Fourthly, the side chains of those residues within 8 Å of the docked ligand were allowed to be flexible. The solutions were ranked based on binding energies and visually examined in *PyMOL*.

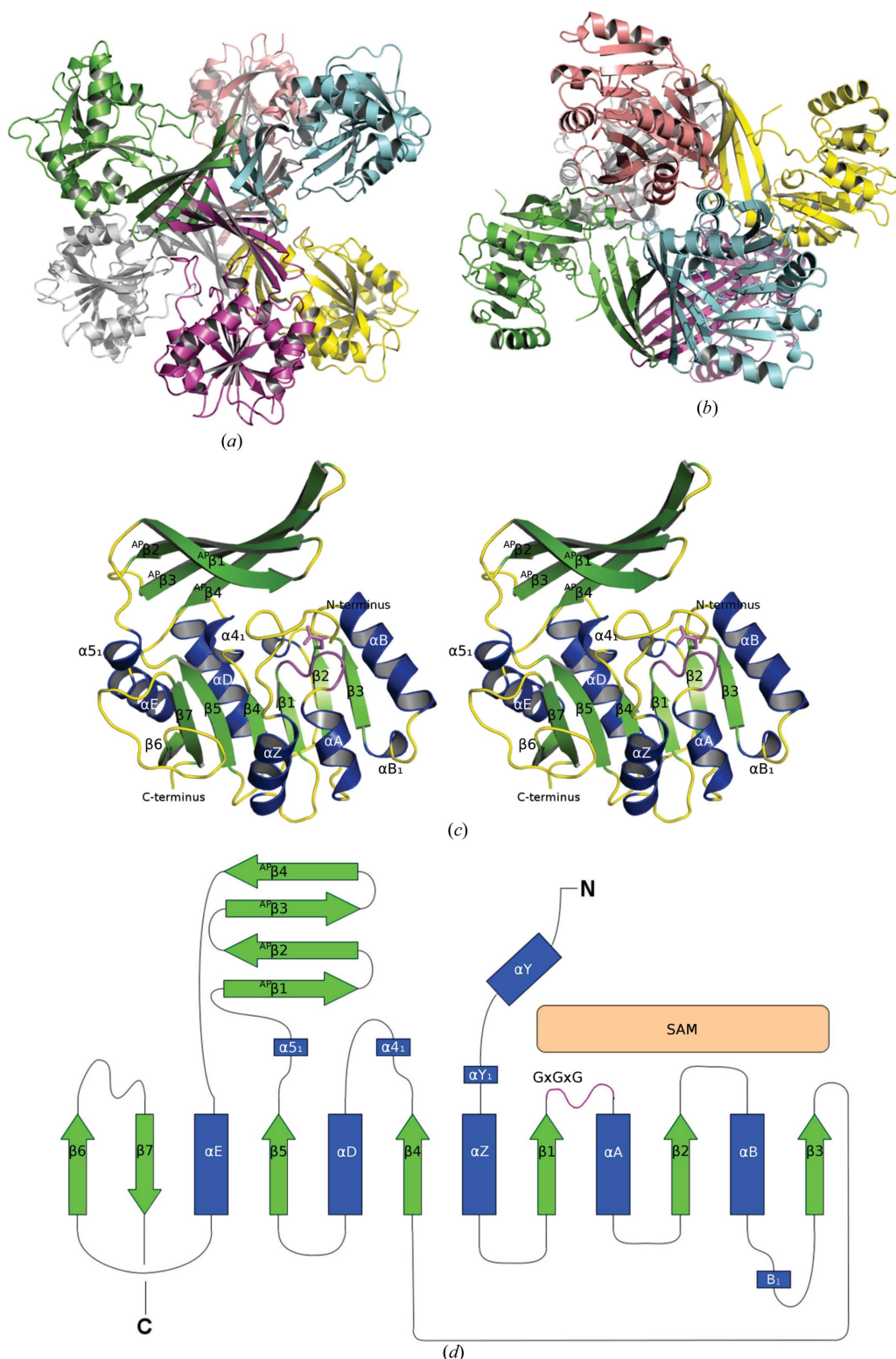
## 2.7. Activity assays

The standard methylation reaction was carried out in 20 mM Tris-HCl pH 8.3 in a total volume of 30 µl. The

**Figure 2**

A weighted  $2F_o - F_c$  electron-density map of the 2.7 Å resolution native structure at  $1.5\sigma$  (which corresponds to  $0.240 \text{ e} \text{ \AA}^{-3}$ ). Side chains can be clearly differentiated in most places and main-chain carbonyl atoms can be distinguished. This view shows the empty cofactor binding site. The conserved GxGxG motif and the acidic residue at the C-terminal end of  $\beta_2$  (Glu71) are coloured magenta.





**Figure 3**

Overall structure of CcbJ. (*a*, *b*) Views of the hexamer in the asymmetric unit looking down (*a*) the noncrystallographic threefold axis and (*b*) the noncrystallographic twofold axis. The view in (*a*) can be transformed into that in (*b*) by rotating  $90^\circ$  about the horizontal axis followed by  $-30^\circ$  about the vertical axis. The individual monomers are coloured separately. (*c*) A stereoview of the CcbJ monomer.  $\alpha$ -Helices are coloured blue,  $\beta$ -strands green and loops yellow. The conserved GxGxG motif and the acidic residue at the C-terminal end of  $\beta 2$  (Glu71) are coloured magenta. These conserved regions also indicate the cofactor binding site. (*d*) A topology diagram of CcbJ. The  $\alpha$ -helices are labelled with capital letters and the  $\beta$ -strands with numbers; the labels match those of catechol *O*-methyltransferase for easy comparison. Short  $3_{10}$ -helices are labelled according to whichever secondary-structural element immediately precedes them; thus,  $\alpha B_1$  immediately follows  $\alpha B$  while  $\alpha 4_1$  immediately follows  $\beta 4$ . The  $\beta$ -strands of the antiparallel  $\beta$ -sheet which forms an active-site cover are labelled  $^{AP}\beta 1-4$  (where AP represents antiparallel).  $\alpha Y$  and the  $3_{10}$ -helix  $\alpha Y_1$  are only visible in the SAH complex. The cofactor binding site is indicated.

reaction mixture contained 3.22 mM *N*-demethylincomycin and 4 mM SAM. The reaction was started by the addition of 20 µg purified CcbJ or mutant and was incubated at 45°C for 2 min. The enzyme was then heat-inactivated (82°C for 2 min) and the amount of lincomycin produced was measured by UPLC (Olšovská *et al.*, 2007) using a BEH C18 column (2.1 × 50 mm internal diameter, particle size 1.7 µm; Waters), with the following mobile phase: solvent *A*, 1 mM ammonium formate pH 9.0; solvent *B*, acetonitrile [76:24(*v:v*)]. The flow rate was 0.5 ml min<sup>-1</sup>, the column temperature was 35°C and the data sample rate was 20 points s<sup>-1</sup> with a filter constant of 0.5. The injection volume was 5 µl, the analysis time was 3.5 min and the UV detection wavelength was 194 nm. Each analysis was followed by a 1 min column-washing step (pure acetonitrile) and a 30 s equilibration step. The results reported in Table 4 are the averages from three independent measurements; the error range covers the 95% confidence interval.

### 3. Results

#### 3.1. Solution and model quality

Native CcbJ (*i.e.* N-terminally His-tagged wild-type CcbJ), its SeMet derivative and its SAH complex all crystallized in space group *C222*<sub>1</sub> with approximately the same unit-cell parameters ( $a \simeq 170$ ,  $b \simeq 244$ ,  $c \simeq 118$  Å). All three crystals appeared to be nearly isomorphous in that meaningful maps could be calculated for each data set using the SeMet structure without resorting to molecular-replacement rotation and translation searches. The asymmetric unit of each crystal contains six chains. These chains are nearly identical, with root-mean-squared deviations over all C<sup>α</sup> atoms in the chain of 0.4 Å for the SeMet derivative, 0.6 Å for the native structure and 0.2 Å for the SAH complex. The overall superposition of all C<sup>α</sup> atoms of chains from all three structures is 0.5 Å. The monomers in the native and SeMet derivatives contain residues 20–253, and the most flexible parts of the protein include residues 40–45, 60–80 and 99–112; the first 26 residues of one chain (chain *B* in the deposited structure) were quite disordered and could only be built reliably beginning from residue 27. In the SAH complex, in contrast, it was possible to build residues 2–253, and the additional residues appear to be involved in forming the SAH binding pocket. In this structure, only the surface loop containing residues 40–45 is highly variable. Representative electron density for the native structure may be seen in Fig. 2; that for the CcbJ–SAH complex structure may be seen in Supplementary Fig. S1.

Structural quality checks using *PROCHECK* (Laskowski *et al.*, 1993) showed that all residues were in the allowed regions of the Ramachandran plot; the somewhat more exacting *MolProbity* (Chen *et al.*, 2010) showed that ten residues in the native structure, four in the SeMet derivative and four in the SAH complex were outside the 99% expected region. In all but one case these residues are in poorly defined loop regions whose geometry was slightly distorted during the refinement process. The exception is Phe117, the  $\varphi$  angle of which is

slightly too narrow. This particular backbone conformation appears to be somewhat stabilized through hydrogen bonds to its backbone N atom from the side chain of Gln149 and to the carbonyl O atom of the preceding Ala116 by both the N atom and the hydroxyl side chain of Thr119. This conformation may not be accidental since, as will be seen later, Phe117 is vital for the proper functioning of the enzyme and most likely interacts with the CcbJ substrate.

#### 3.2. Overall structure and topology

The overall structure of CcbJ can be seen in Fig. 3 along with a topology diagram. The final models have *R*-factor and *R*<sub>free</sub> values of 0.181 and 0.224 for the native structure, 0.225 and 0.248 for the SeMet derivative and 0.183 and 0.221 for the SAH complex, respectively. The six chains in the asymmetric unit form a hexamer which is arranged as a dimer of trimers with approximate 32 point symmetry (Figs. 3*a* and 3*b*; see also Supplementary Fig. S2). Electron micrographs, gel filtration and native electrophoresis all indicate that this hexamer is most likely to exist in solution as well (Najmanová *et al.*, 2013).

CcbJ has a parallel  $\alpha/\beta$  structure in which an open parallel  $\beta$ -sheet is flanked on both sides by  $\alpha$ -helices. The overall topology of CcbJ matches that observed in other class I SAM-dependent methyltransferases. It also possesses the characteristic markers of this class (Martin & McMillan, 2002), *viz.* a GxGxG motif at the end of  $\beta$ 1 and an acidic residue, Glu71, at the end of  $\beta$ 2 (Figs. 3*c* and 3*d*). It has a few additions to the core fold which are characteristic of small-molecule methyltransferases. These include an additional  $\alpha$ -helix at the N-terminus, a somewhat extended loop between  $\beta$ 6 and  $\beta$ 7, and an active-site cover which is inserted between  $\beta$ 5 and  $\alpha$ E. This cover takes the form of a four-stranded antiparallel  $\beta$ -sheet containing antiparallel (AP) strands <sup>AP</sup> $\beta$ 1–<sup>AP</sup> $\beta$ 4. There is no C-terminal extension and no additional insertions into the core fold. Helix  $\alpha$ C is absent in CcbJ, its place being taken by an extended loop (Fig. 3*d*), but this is not unusual for this class.

#### 3.3. Subunit interactions

The CcbJ hexamer appears to be arranged as a dimer of trimers. The trimer interface is formed by hydrophobic packing between the side chains on the outside of the antiparallel  $\beta$ -sheet which forms the active-site cover (Fig. 4*a*). This packing exhibits nearly exact threefold symmetry (average rotation of 120 ± 2°), and the presence of such a high concentration of hydrophobic residues on what would otherwise be a solvent-exposed area strongly argues against the presence of individual monomers in solution. Analysis of the native and SeMet CcbJ structures using *Protein Interfaces, Surfaces and Assemblies* (PISA; Krissinel & Henrick, 2007) shows that the assembly generated by this interface buries ~4000 Å<sup>2</sup> of accessible surface area and has a solvation free-energy gain  $\Delta G^{\text{int}}$  of -21.0 kcal mol<sup>-1</sup> and a free energy of dissociation  $\Delta G^{\text{diss}}$  of 8.3 kcal mol<sup>-1</sup> (it should be noted that these results were obtained using only the protein molecules

themselves in the absence of additional sulfate anions or ethylene glycol molecules).

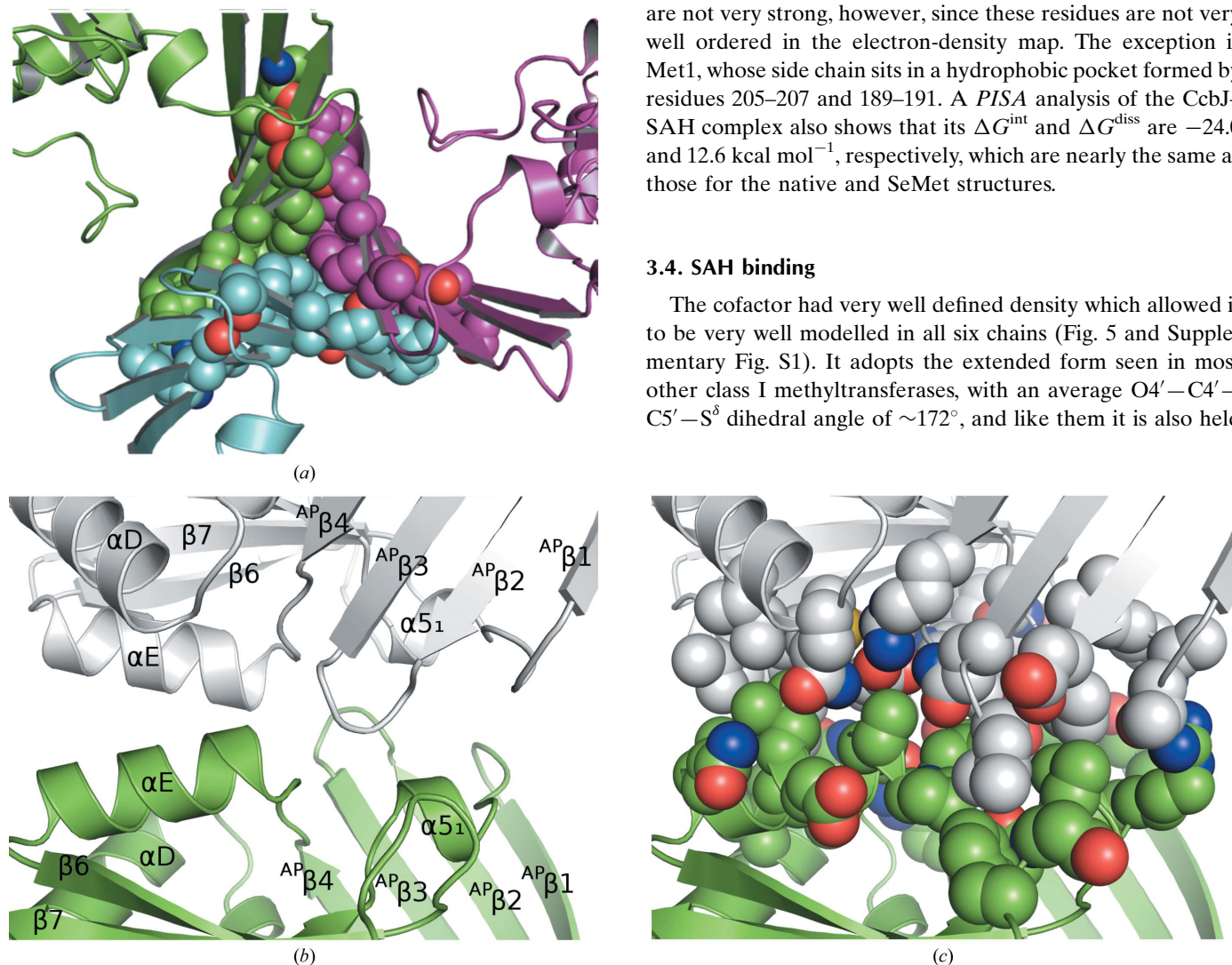
The interface between the two trimers is larger, with  $\sim 10\,000\text{ \AA}^2$  of accessible surface buried when two trimers come together to form the hexamer, and it may be somewhat stronger since  $\Delta G^{\text{int}}$  for the entire hexamer is approximately  $-153\text{ kcal mol}^{-1}$  and  $\Delta G^{\text{diss}}$  is  $64.2\text{ kcal mol}^{-1}$ . The interface between any two of the individual monomers is almost exactly twofold, with an average rotation of  $179 \pm 2^\circ$  (Figs. 4*b*, 4*c* and Supporting Information). This interface is centred around the interaction between  $\alpha E$  and its counterpart in the second monomer. These two helices pack against each other in a roughly antiparallel manner, while residues from  $\alpha D$  fill in grooves at the edge. The loop connecting  $\beta 5$  to the antiparallel  $\beta$ -sheet also interacts with its counterpart from the adjacent

monomer in a similar way. The residues at the edge of the antiparallel  $\beta$ -sheet interact both with the residues at the N-terminal edge of  $\alpha E$  from the adjacent monomer, and with residues 239–241. As will be seen later, the residues between  $\beta 5$  and the active-site cover are in a position to potentially interact with bound substrate. These interactions give a hexamer with the overall shape of an oblate spheroid with a diameter of about  $72\text{ \AA}$  along the short (threefold) axis and  $100\text{ \AA}$  along each of the long (twofold) axes.

These subunit interactions change very little in the SAH complex. The greatest change is that binding of the demethylated cofactor has allowed the N-terminal 19 residues to become ordered and some of these residues now participate in the trimer interface (see Supplementary Fig. S1). Specifically, Met1, Arg2, Asn3 and Gln12 are now within van der Waals contact distance of residues 168–172, which form the first  $\beta$ -strand of the antiparallel  $\beta$ -sheet. These new interactions are not very strong, however, since these residues are not very well ordered in the electron-density map. The exception is Met1, whose side chain sits in a hydrophobic pocket formed by residues 205–207 and 189–191. A PISA analysis of the CcbJ–SAH complex also shows that its  $\Delta G^{\text{int}}$  and  $\Delta G^{\text{diss}}$  are  $-24.0$  and  $12.6\text{ kcal mol}^{-1}$ , respectively, which are nearly the same as those for the native and SeMet structures.

### 3.4. SAH binding

The cofactor had very well defined density which allowed it to be very well modelled in all six chains (Fig. 5 and Supplementary Fig. S1). It adopts the extended form seen in most other class I methyltransferases, with an average  $\text{O4}'\text{—C4}'\text{—C5}'\text{—S}^{\delta}$  dihedral angle of  $\sim 172^\circ$ , and like them it is also held



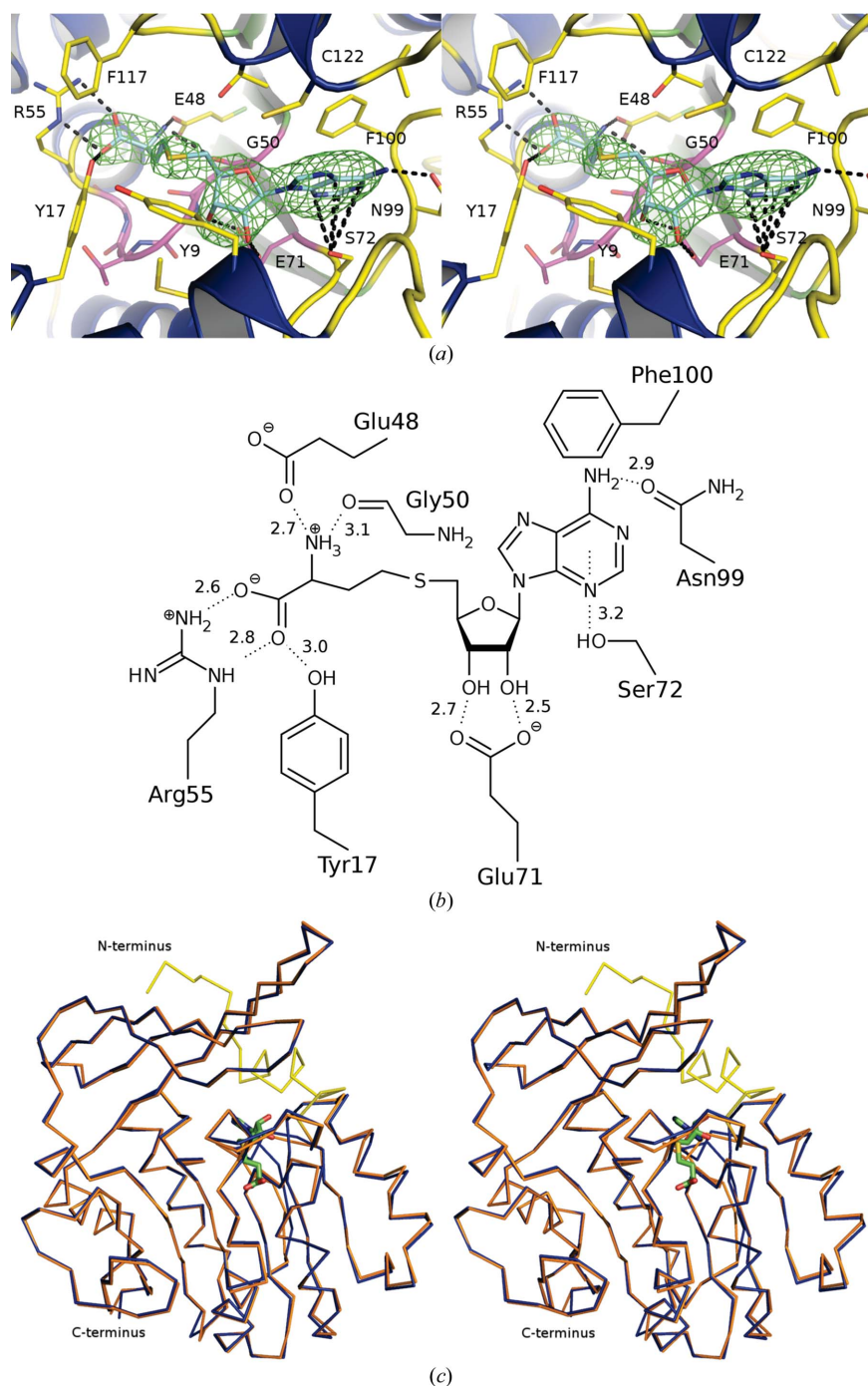
**Figure 4**

Subunit interfaces. (a) The interface between the three chains which form the trimer arises primarily from hydrophobic packing of the side chains on the outside surface of the antiparallel  $\beta$ -sheet (that is, the surface which would face the solvent if the monomers were separate). The picture shows these side chains as van der Waals radius spheres. (b) The dimer interface viewed perpendicular to the twofold axis with the secondary-structural elements labelled. Three of these bring the trimers together to form the hexamer. The two interacting chains are shown in green and grey. The interface is primarily arranged around helix  $\alpha E$  with contributions from the residues belonging to the  $\alpha 5_1$   $3_{10}$ -helix and those belonging to the hairpin loops joining the strands of the antiparallel  $\beta$ -sheet. (c) The same view as in (b), but showing those side chains which pack together to form the interface.



in place by an extensive series of hydrogen bonds and hydrophobic interactions.

The binding pocket contains a few unusual features, but it is not dissimilar to those seen in other methyltransferases (see



**Figure 5**

The SAH binding site. (a) The binding site of the *S*-adenosyl-L-homocysteine inhibitor, showing those residues which interact with it; the most important of them are labelled. The conserved GxGxG and acidic residues are coloured magenta. The electron-density map is a positive weighted  $F_o - F_c$  difference map shown at a contour level of  $4\sigma$  (corresponding to  $0.122 \text{ e } \text{Å}^{-3}$ ). (b) A schematic view of the cofactor binding site, showing SAH (at the centre) and the residues which participate in binding to it. Dotted lines designate hydrogen bonds and the distances are given in angstroms. This figure was drawn using *MarvinSketch* 6.1.3 from ChemAxion. (c) A  $C^\alpha$  trace showing the superposition of the native structure (blue) with the CcbJ-SAH complex structure (orange). Binding of the SAH inhibitor causes the N-terminal 19 residues to become ordered. These ordered residues, shown in yellow, form an  $\alpha$ -helix and a short  $3_{10}$ -helix.

Fig. 5b; Fauman *et al.*, 1999; Schubert *et al.*, 2003; Liscombe *et al.*, 2012). The carboxylic acid group of the homocysteine portion is hydrogen-bonded to the guanidinium group of Arg55, which precedes the GxGxG motif, and the hydroxyl group of Tyr17, while the amine N atom appears to be hydrogen-bonded to the side chain of Glu48 and the carbonyl O atom of Gly50, one of the glycines of the GxGxG motif. The O2' and O3' hydroxyl groups of the ribose sugar are hydrogen-bonded to the side chain of Glu71, the conserved acidic residue at the end of  $\beta 2$ . The adenine ring is sandwiched between the side chains of Phe100, which appears to base-stack with it, and Ser72. Ser72 may form an uncommon  $\text{OH} \cdots \pi$  hydrogen bond similar to that which appears in the *Thermus thermophilus* TTHA1280-SAH complex (Pioszak *et al.*, 2005; PDB entry 2cww). The N6 atom is within hydrogen-bonding distance of the carbonyl O atom of the Asn99 side chain, while the N1 and N3 atoms appear to be within hydrogen-bonding distance of the backbone amide N atoms of Phe100 and Ser72, respectively. Binding of SAH causes the N-terminal 19 residues, which are disordered in the native and SeMet-derivative structures, to fold over the top of the SAM binding cavity and residues 8–13 to fold into an  $\alpha$ -helix, while residues 14–19 form a quasi-helix of hydrogen-bonded turns which do not have the correct geometry to form a helix (Fig. 5c). If SAM is modelled in the cofactor binding site by adding a  $C^\epsilon$  atom to the  $S^\delta$  of the inhibitor, the  $C^\epsilon$  atom, which belongs to the transferred methyl group, is in van der Waals contact with Tyr9, Tyr17 and Phe117.

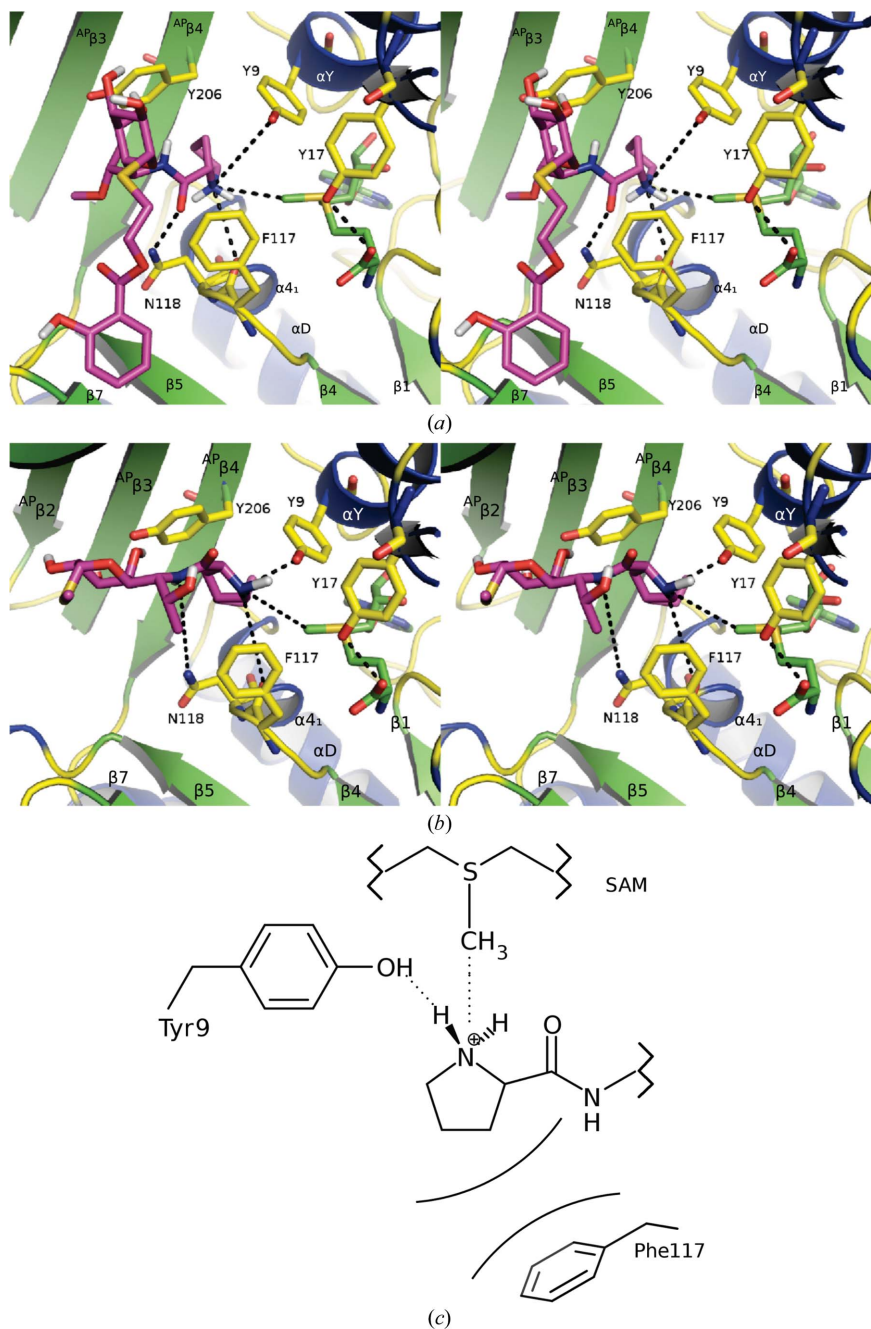
In the probable substrate-binding sites of five of the six monomers, a continuous patch of density was found which was modelled as glycerol (see Supplementary Fig. S3 for an example). In four of these cases, this glycerol was closer than  $5.6 \text{ \AA}$  to the  $S^\delta$  atom of SAH, which is the minimum van der Waals distance that we would expect to observe for an  $S^\delta - C^\epsilon \cdots \text{HO}$  alignment. In one of these cases, the glycerol OH and SAH  $S^\delta$  were only  $3.6 \text{ \AA}$  apart, which is the van der Waals distance of these two groups (Li & Nussinov, 1998). In addition, a blob of density was observed between chains *D* and *B* which was built and refined as a molecule of HEPES buffer.

3.5. Structural homologies

The protein structure-comparison service *PDBFold* (Krissinel & Henrick, 2004) was used to search the Protein Data Bank for structures similar to CcbJ with the monomer,

dimer and trimer as search models. CcbJ is most similar to other SAM-dependent methyltransferases. In particular, it is most like those of the CAC2371-like family (SCOP 117688), which has a very similar overall fold to the glycine *N*-methyltransferase (SCOP 53348; GNMT) and mRNA cap (guanine N-7) methyltransferase (SCOP 102560) enzymes; in fact, both of these enzymes also appeared in the search results. On the other hand, the sequence identity of aligned residues was rather low, with the best alignment (204 aligned residues) having a sequence identity of only 23%, while the lowest (159 residues) had a sequence identity of only 17% (the best structure match had the lowest identity of all, at only 10%). This is not unexpected and is actually characteristic of SAM-dependent methyltransferases. It is well known that the only conserved part of this motif is the GxGxG motif found between  $\beta 1$  and  $\alpha A$  and the acidic residue normally located at the C-terminal end of  $\beta 2$  (Martin & McMillan, 2002). These motifs are indicated in Fig. 3(c).

The structure with the highest similarity to CcbJ is that of an unknown SAM-dependent methyltransferase from *Clostridium acetobutylicum* ATCC 824 (*Q*-score 0.5344, *Z*-score 12.4, r.m.s.d. 1.87 Å; PDB entry 1y8c; New York SGX Research Center for Structural Genomics, unpublished work). Supplementary Fig. S4 shows an overlap of this structure with native CcbJ. Despite having the best overall structural similarity, this protein has the lowest sequence similarity among matched residue pairs and the residues forming the cofactor-binding pocket are completely different between the two proteins. A more interesting match is to an uncharacterized methyltransferase from *Pyrococcus horikoshii* OT3 (*Q*-score 0.4908, *Z*-score 11.9, r.m.s.d. 2.07 Å; PDB entry 1wzn; RIKEN Structural Genomics/Proteomics Initiative, unpublished work). Although it has a somewhat poorer structural match with the CcbJ monomer, this protein does share its quaternary structure. Like CcbJ, it appears to be a hexamer which is built up from a dimer of trimers; the trimers are also held together by hydrophobic interactions between the side chains of an antiparallel  $\beta$ -sheet which sits over the top of the probable active site, and the trimer-trimer interactions are also similar and are centred around  $\alpha E$ . More interestingly, this protein also has equivalents of Tyr9, Tyr17 and Phe117, three residues which may be



**Figure 6** Substrate binding to CcbJ. (a) A stereoview showing the lowest energy docking solution for *N*-demethylcelesticetin to a CcbJ-SAM complex most likely to favour the methyltransfer reaction. The four most important residues are shown along with Tyr206. (b) A stereoview showing the lowest energy docking solution for *N*-demethylincomycin to a CcbJ-SAM complex most likely to favour the methyltransfer reaction. (c) A schematic view of the area around the reaction center for the lowest energy docking solution of protonated *N*-demethylcelesticetin (in the centre) and a CcbJ-SAM complex. The hydroxyl group of Tyr9 is within hydrogen-bonding distance of the substrate, while Phe117 makes van der Waals contacts with it. In this solution, the N1' target atom is only 3.7 Å from the transferred methyl group of the cofactor. This figure was drawn using *MarvinSketch* 6.1.3 from ChemAxon.



important for the catalytic activity of the enzyme, as well as Arg55 (which is conserved in space, but not in sequence; in the *P. horikoshii* methyltransferase the guanidinium side chain is contributed by Arg19; see Supplementary Fig. S5). In most other respects, however, the two enzymes differ, with completely differently shaped active sites and cofactor-binding sites.

Superimposing all of the matching structures shows that their nucleotide-binding sites superimpose quite well and that, with the exception of the GNMT structure with PDB code 1xva (which is most likely not to be the physiologically correct conformation; Huang *et al.*, 2000), the SAM or SAH moieties of all of the complex structures superimpose closely and have the same overall conformation.

### 3.6. Computational docking

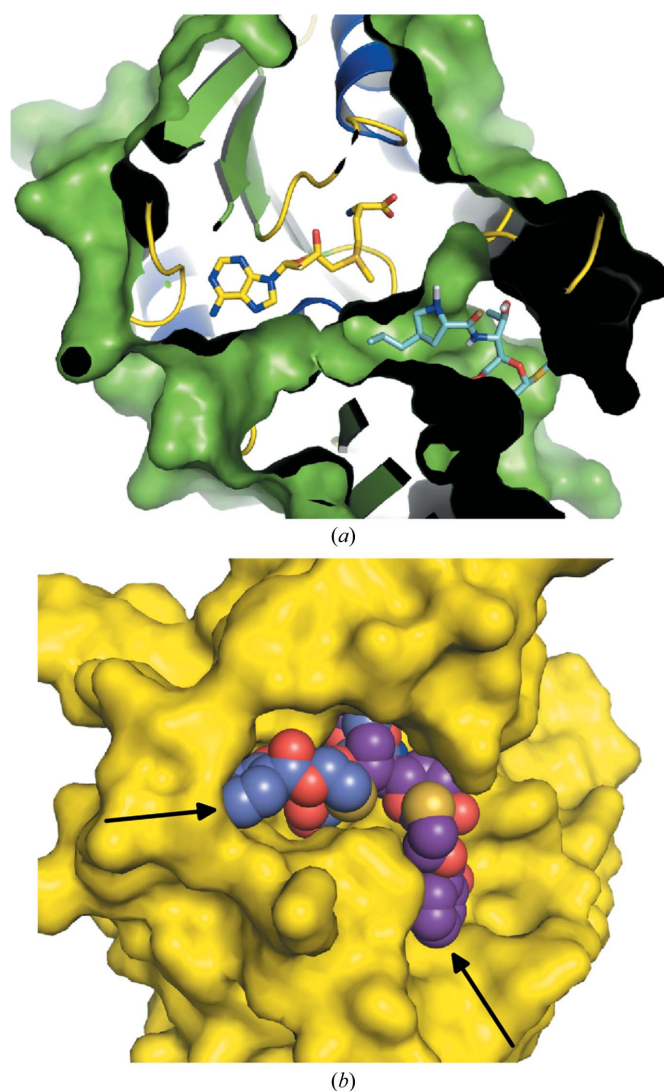
Unfortunately, neither celesticetin nor any of its demethylated or desalicylated forms are available commercially; therefore, no compounds were available for co-crystallization experiments. Consequently, we carried out computational docking of CcbJ–SAH and CcbJ–SAM complexes with celesticetin, *N*-demethylcelesticetin, lincomycin and *N*-demethylincomycin using *AutoDock Vina* (Trott & Olson, 2010). Since the true physiological substrate of CcbJ is unknown, we also carried out docking simulations using *O*-demethylcelesticetin, *N,O*-didemethylcelesticetin, desalicyetin, *N*-demethylsalicyetin, *O*-demethylsalicyetin and *N,O*-didemethylsalicyetin. At physiological pH as well as at the pH used in our assays, all substrates are expected to be protonated on the target atom (the N1' atom of the propylproline moiety); consequently, all docking experiments were carried out using both protonated and unprotonated forms. The crystal structure of lincomycin reported by Rajeswaran & Srikrishnan (2004) has an *R* configuration at the target N1' atom; as a result, the H atom which would lead to an *R* configuration was removed from the *N*-demethylated substrates.

Supplementary Table S2 summarizes the docking energies of each compound. The celesticetin variant with the best docking energies was protonated *N*-demethylcelesticetin, followed closely by *N*-demethylcelesticetin. Of the lincomycin variants, both protonated and deprotonated *N*-demethylincomycin were equally good. The docking conformations of protonated *N*-demethylcelesticetin and deprotonated *N*-demethylincomycin were therefore examined in greater detail and can be seen in Figs. 6(a) and 6(b). The clustering of the top six solutions for these two ligands can be seen in Supplementary Fig. S6.

All class I methyltransferases reported to date have a substrate-binding site located over the ends of  $\beta$ -strands 5–7. Both ligands docked into this predicted binding cleft. Celesticetin and *N*-demethylcelesticetin docked better than lincomycin and *N*-demethylincomycin. The docking energies for doubly protonated *N*-demethylated celesticetin were  $-8.1$  to  $-8.8$  kcal mol<sup>-1</sup>, and those for the doubly protonated *N*-demethylated lincomycin were  $-6.6$  to  $-7.7$  kcal mol<sup>-1</sup>; the

energies for the singly protonated forms were comparable, although slightly higher for *N*-demethylcelesticetin and somewhat lower for *N*-demethylincomycin. The energies for the products celesticetin and lincomycin (protonated forms) were  $-7.4$  to  $-8.7$  kcal mol<sup>-1</sup> and  $-7.3$  to  $-8.0$  kcal mol<sup>-1</sup>, respectively; as before, the energies of the deprotonated forms were comparable. Those complexes which place the substrate target atom closest to the C<sup>e</sup> methyl group of SAM had docking energies of  $-8.5$  to  $-8.7$  kcal mol<sup>-1</sup> for *N*-demethylcelesticetin and  $-7.3$  kcal mol<sup>-1</sup> for *N*-demethylincomycin.

The substrate-binding pocket of CcbJ is quite large and has a clear funnel shape: a very wide cleft on the surface narrows almost to a point focused on the S <sup>$\delta$</sup>  atom of the SAH inhibitor (or on the transferred C<sup>e</sup> atom of the SAM cofactor were it bound). Interestingly, there is also a deep narrow pocket



**Figure 7**  
The CcbJ active-site opening. (a) A cutaway view of the enzyme showing the shape of the active-site pocket. The propylproline tail of *N*-demethylincomycin shows the small side passage. (b) An exterior view of the CcbJ active site with two of the lower energy docking conformations of the substrate shown (blue and purple van der Waals spheres). There are at least two possible locations for the salicylic acid tail of an *N*-demethylcelesticetin substrate (indicated by arrows).

which opens out into a channel in four of the six monomers, which begins just in front of the S<sup>δ</sup> atom, passes through an opening flanked by Tyr9 and Tyr206, and continues through a passage made up of <sup>AP</sup>β4, αY and α4<sub>1</sub>. One side of this channel also passes over the ribose sugar and the edge of the adenine ring of the cofactor. At its narrowest, this channel is ~5.5 Å across. It is too narrow to be completely solvent-accessible, but might be large enough to admit a single water molecule or a hydrocarbon chain.

According to the docking results, the following residues are within 8.0 Å of the docked substrate and are therefore likely to be involved in enzyme–substrate contacts: Tyr9, Tyr17, Phe117, Asn118, Phe121, Leu151, Gln156, Arg157, Leu158, Thr163, Val167, His175, Glu177, Ser179, His181, Leu188, His192, Tyr206, Leu208, Arg244 and Tyr245. Of these, Tyr9, Tyr17 and Phe117 are also in contact with the cofactor and may therefore be important for the catalytic reaction (see §4). In those conformations most likely to be suitable for the methyl-transfer reaction to occur, the target N1' atom of the substrate is within hydrogen-bonding distance of the hydroxyl group of Tyr9 and in van der Waals contact with Phe117; frequently, Asn118 is close enough to form a hydrogen bond to the amide bond connecting the amino-acid and sugar moieties of the substrate (Fig. 6c).

The active-site cover has a number of charged and polar residues on its inner surface, including Thr163, Glu168, His175, Glu177, Ser179, His192 and Tyr206, while the opposite side includes Phe117, Leu151, Leu158, Leu188 and Tyr245. This distribution should make the binding cavity somewhat more negatively charged on one side than the other. Perhaps as a consequence of this, the docking simulations tended to orient the hydroxyl groups of the substrate sugar ring towards the active-site cover rather than towards the opposing loop.

The *N*-demethylincomycin docking results are similar to those for *N*-demethylcelesticetin. Notably, they also place the N1' target atom within hydrogen-bonding distance of the hydroxyl group of Tyr9, but in addition they place the three atoms of the additional propyl group attached to the proline moiety of lincomycin into the pocket or channel observed close to the reaction centre (Fig. 7a).

### 3.7. Point mutants

The SAH complex structure and the docking complexes suggested that the residues Tyr9, Tyr17 and Phe117 might be important for the enzymatic reaction. Y9F, Y17F and F117G mutants were therefore prepared to study the effects of these changes on the catalytic activity of the enzyme. The Y9F and Y17F mutations were chosen to abolish the hydrogen-bonding capabilities of these two residues, while the F117G mutation was chosen to eliminate any van der Waals contact between this residue and the substrate. The overexpression and purification procedures for the native and mutant CcbJ proteins were the same. The yield of mutant proteins was comparable with that of the native and, as Supplementary Fig. S7 shows, all mutants showed generally the same pattern as the native on blue native polyacrylamide gel electrophoresis; this indicates

that the oligomeric state of the mutants is likely to be the same as that of the native CcbJ.

The methylation activity of the mutant recombinant proteins was compared with that of native CcbJ *in vitro* using a standard methylation reaction. After 2 min of incubation at 45°C with *N*-demethylincomycin, the concentration of the lincomycin product was measured by UPLC (Olšovská *et al.*, 2007). The results of these assays can be seen in Table 4, which reports the amount of lincomycin produced in nanograms per microgram of CcbJ. The lincomycin concentration produced by the F117G and Y9F mutants was very close to the background level, while the Y17F mutant managed to produce a level almost half that of the wild-type enzyme.

## 4. Discussion

We report here the structures of the CcbJ methyltransferase from *S. caelestis* determined by MAD phasing to 3.0 Å resolution, its native form to 2.7 Å resolution and the CcbJ–SAH complex to 2.9 Å resolution. We also describe the results of biochemical assays carried out using point mutants based on these structures along with those of docking simulations performed using these structures. Our results allow us to propose a plausible model for its substrate binding along with a likely description of its reaction mechanism.

The substrate-binding cavity of CcbJ is cone-shaped, with a wide opening narrowing down to a focus near the S<sup>δ</sup> atom of the inhibitor. If the C<sup>ε</sup> atom of the SAM cofactor is modelled in, it sits near the focus of the whole binding cavity. This shape immediately suggests that CcbJ may be able to accommodate several possible salicylic acid derivatives attached to the primary hydroxyl group of the precursor desalicytin. Computational docking experiments showed that at least two different positions are available for accommodating the salicylic acid moiety at the opening of the active site (Fig. 7b).

Moving towards the reactive centre, one side of the binding cleft, the side formed by the antiparallel β-sheet active-site cover, is notably more negatively charged than the opposite face as a result of the presence of a number of charged or polar residues (Supplementary Fig. S8). This asymmetry is most likely to promote the orientation of the substrate so that the hydroxyl groups of the sugar ring are oriented towards this region of the active-site cover. Closer in towards the reactive centre, the active site becomes more hydrophobic as a result of the presence of Tyr9, Tyr17, Phe117 and Tyr206. Leading off from the reactive site is a small pocket which opens into a narrow channel in four of the six monomers, which is just wide enough to accommodate an unsaturated hydrocarbon chain and would account for the ability of CcbJ to process both celesticetin, with an unmodified proline moiety, and lincomycin, which has a propylproline moiety. Docking studies using *N*-demethylincomycin placed the propyl group of the lincomycin propylproline moiety into this channel in the lowest energy conformations (Fig. 7a). This channel, along with the wide active-site opening, suggests that CcbJ may be able to methylate a variety of semi-synthetic lincosamide antibiotics, including those with a very different moiety from

the salicylic acid group and an addition to the proline ring of longer than three C atoms. Indeed, we have previously shown that the lincomycin-biosynthetic pathway, and therefore the CcbJ homologue LmbJ, is able to produce several lincomycin derivatives with extended alkyl chains attached to the proline moiety (Ulanova *et al.*, 2010).

At the reaction centre, three aromatic residues, Tyr9, Tyr17 and Phe117, surround the location where the reaction is expected to occur; when SAM is modelled into the cofactor binding site in place of SAH, these three residues are within 4.0 Å of the transferable methyl group. To determine whether these residues play a role in the methyltransferase reaction, three point mutants were made: Y9F, Y17F and F117G. Y9F and F117G reduced the production of lincomycin to the background level, while Y17F reduced it to less than half of the normal wild-type level. The work of Hegazi *et al.* (1976, 1979) and Woodard *et al.* (1980) established that catechol *O*-methyltransferase is most likely to carry out its methyltransfer reaction through an S<sub>N</sub>2 mechanism, and this mechanism is generally thought to apply to all class I methyltransferases with nucleophilic N, O and S target atoms (Fauman *et al.*, 1999; Schubert *et al.*, 2003). For the reaction to occur, therefore, the acceptor atom of the substrate must be in a straight line with the C<sup>ε</sup> and S<sup>δ</sup> atoms of the SAM cofactor. These three residues are most likely to contribute to forming this alignment, and Tyr9 and Phe117 appear to be vital to it.

Tyr9 is most likely to act by forming a hydrogen bond to the target N1' atom of the substrate. Under physiological conditions, as well as those of our biochemical assays, the substrate would be likely to be doubly protonated on the target N atom. In order to produce an *R* configuration of the target atom, which appears to be the configuration of lincomycin seen in its crystal structure (Rajeswaran & Srikrishnan, 2004), the H atom in the N1'–H···OH hydrogen bond to Tyr9 must be removed. Therefore, Tyr9 most likely fulfills a dual function as a general base to extract a blocking H atom and to help position the substrate in the correct conformation. In support of this, the Y9F mutant exhibits almost no reactivity and a docking study using a Y9F mutant CcbJ–SAM complex structure as the receptor does not reproduce the conformations seen in Fig. 6 for either substrate (the substrate always docks so that the target atom faces away from the C<sup>ε</sup> atom of the cofactor). Furthermore, the hydroxyl group of Tyr9 is potentially solvent-accessible, which would make removal of this extra H atom easier.

Formation of the correct substrate conformation would also be supported by Phe117. The aromatic side chain forms one side of a small pocket just in front of the cofactor which directs the propylproline ring of the substrate towards the reactive centre. The docked conformation seen in Fig. 6(a) indicates that the main-chain carbonyl of this residue may also play a role: in the lowest energy docked conformation, it is within hydrogen-bonding distance of the target atom through the H atom which would be retained in the *R* conformation of the product. This might account for the slightly strained backbone conformation of this residue observed in the native structure. A docking study using an F117G mutant as the receptor shows

that in the absence of the Phe117 side chain the sugar moiety of the substrate tends to fill the hole where the phenyl side chain had been, while the target N1' atom remains within hydrogen-bonding distance of the Tyr9 hydroxyl group. This causes the substrate to rotate so that the C atom adjacent to the target N1', C5', is the closest atom to the transferable methyl group.

These results indicate that both of these residues are important for correctly orienting the substrate for the methyltransfer reaction to take place. The substrate binds so that the propylproline amino-acid moiety is positioned between the side chains of Phe117 and Tyr9. Tyr9 assists the methyltransfer reaction by extracting an H atom from the target atom and helping to position the substrate. Phe117 on the other side of the reaction centre also helps to position the substrate moiety correctly, while its carbonyl O atom may hydrogen-bond to the substrate target atom, anchoring it in place and further assisting in the development of a reactive conformation. These two interactions play a crucial role in properly positioning the substrate so that the lone pairs of the N1' acceptor atom are properly aligned with the C<sup>ε</sup> atom of the cofactor. If either of these interactions is disrupted, the substrate will be likely to not align properly and the reaction will not take place.

Our docking simulations show that the torsion angle around the bond between the C<sup>α</sup> and C atoms of the propylproline moiety frequently rotates away from the conformation observed in the original lincomycin crystal structure. It is tempting to conclude that the interactions of Tyr9 and Phe117 with the substrate cause rotation around this bond and that this rotation is an important part of the process which aligns the substrate acceptor atom with the methyl donor. If this is correct, then the barrier to rotation around this bond must be on roughly the same level as a hydrogen bond. The energy of a single hydrogen bond within a protein is normally taken to be 1.5 kcal mol<sup>-1</sup>, although higher values are possible depending on local conditions (Anslyn & Dougherty, 2006). Calculating the barrier of rotation using *Avogadro* v.1.1.1 (Hanwell *et al.*, 2012) with the MMFF94 forcefield (Halgren, 1996) shows that an angle of ~190° has the lowest energy for all substrates and that this torsion angle can rotate by 20–30° before exceeding this value. A rotation of 30° would make this torsion angle 220°, which is approximately the angle observed in the docking studies. This prediction is consistent with the idea that rotation around the propylproline C<sup>α</sup>–C bond is part of the methyl-transfer process. We are unaware of any similar mechanisms which may have been proposed for other methyltransferases, however, and given the lack of detailed conformational information this proposal must remain speculative.

The role of Tyr17 is somewhat more obscure. It most likely does not interact directly with the substrate using its hydroxyl group because this interaction is blocked by the side chain of Phe117. What it does do is form a hydrogen bond to the carboxylic acid group of the cofactor. In the absence of this hydroxyl group, the methionine moiety of the cofactor is possibly less well secured, potentially allowing the C<sup>ε</sup> atom to



be less well aligned with the substrate, thereby slowing, but not halting, the reaction.

As seen earlier, residues 151–162, which connect  $\beta 5$  to the active-site cover, are in van der Waals contact with their counterparts in an adjacent subunit of the hexamer. Several of these residues, Leu151, Gln156, Arg157 and Leu158 in particular, are within 8.0 Å of the substrate in its lowest energy docked conformation. Furthermore, because they lie at the opening of the substrate-binding pocket, the substrate would need to interact with them at least transiently when binding. This implies that the activity of one CcbJ monomer may be dependent on the state of one or more of its neighbours. What exactly the nature of this interaction is cannot be inferred based on the available information, however.

As noted earlier, the true physiological substrate of CcbJ is presently unknown. Our docking simulations suggest that the best substrate for CcbJ is likely to be *N*-demethylcelesticetin, but they do not rule out the possibility that any of the other molecules examined are also substrates. For example, *N*-demethylincomycin is known to serve as a substrate for CcbJ (Najmanová *et al.*, 2013), and yet its docking energy is nearly 1.5 kcal mol<sup>-1</sup> worse than that of *N*-demethylcelesticetin. Going by docking energies alone, therefore, this also suggests that *N,O*-didemethylcelesticetin might also be a substrate of CcbJ. Furthermore, although the docking energies of all the desalicyetin variants were higher still, docking conformations likely to favour methyl transfer were still found, again suggesting that these molecules may plausibly be substrates for CcbJ.

Desalicyetin and three possible celesticetin intermediates with acetyl, isobutyryl and anthranilyl groups in place of the salicylate group have been recovered from fermentation mixtures of *S. caelestis* (Wright, 1983). There are two possible interpretations for the presence of these side products in the fermentation mixture: either the salicylate unit is gradually synthesized following *N*-methylation of the amino-acid moiety, or *N*-methylation can occur with any salicylate derivative attached. Our structural and docking results for CcbJ favour the latter.

The orthologous protein LmbJ from lincomycin biosynthesis (Koběřská *et al.*, 2008; GenBank accession No. ABX00606.1) shows 59% sequence identity to CcbJ (GenBank accession No. ADB92558), suggesting that they have only relatively recently diverged. As can be seen in Supplementary Fig. S9, all three of the catalytically important residues explored here for CcbJ are also present in LmbJ, suggesting that LmbJ should function in the same way as CcbJ. In addition, most of the CcbJ residues which were predicted to be involved in enzyme–substrate contacts according to the docking results are also present in LmbJ. Only three of the predicted contact residues in CcbJ are very different in LmbJ. Gln156, Arg157 and Thr163 in CcbJ are Lys150, Ser151 and Asp157, respectively, in LmbJ. These residues are found at the outer edge of the active-site opening and would be most likely to be involved in contacts with the salicylate group. These changes, therefore, are most likely to reflect the differences in the structures of the natural substrates.

#### 4.1. Summary

In summary, we have overexpressed, purified and determined the structures of CcbJ by MAD phasing at 3.0 Å resolution, its native form at 2.7 Å resolution and its complex with SAH at 2.9 Å resolution. Using these structures, we constructed three point mutants and carried out a series of docking simulations to produce substrate–SAM–CcbJ and product–SAH–CcbJ complexes in order to determine the likely mechanism of this enzyme. Our results show that CcbJ has a very wide and rather nonspecific active site which allows it to accommodate a variety of substrates. They also allow us to infer that three residues appear to be especially important for the methyltransferase reaction: Tyr9, Tyr17 and Phe117. Tyr9, in combination with Phe117, forms a pocket which accommodates the propylproline moiety of the substrate and helps to properly position it for methyl transfer. It is possible that rotation about the C<sup>α</sup>–C bond of the substrate is part of the process by which the lone pair on the substrate target atom is aligned with the cofactor C<sup>ε</sup> group during methyl transfer. Tyr9 may also contribute by extracting an additional H atom which is expected to be present under physiological conditions and which would prevent the reaction from occurring. Phe117 could also contribute by hydrogen-bonding to the substrate target atom using its carbonyl O atom. Tyr17 probably helps to hold the rest of the methionine moiety of the cofactor in place so that the methyl-transfer reaction can occur. Questions not addressed here, and possible targets of future study, include the dependence, if any, of any given CcbJ monomer on the states of the other five in the hexamer which is likely to be the soluble form and the interactions between CcbJ and the other members of the celesticetin-biosynthetic pathway.

This research was supported by VEGA 2/0122/11 and MVTS 1520 in Slovakia and Ministry of Education, Youth and Sports of the Czech Republic projects IMPULS No. CZ.1.07/2.3.00/30.0003 and BIOCEV – Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University No. CZ.1.05/1.1.00/02.0109 from the European Regional Development Fund in the Czech Republic. We would like to thank Björn Sjöblom and Kristina Djinović-Carugo of the Max F. Perutz Laboratories of the University of Vienna for allowing us the use of their X-ray diffractometer for initial crystal screening and for helpful conversations. We would also like to thank the staff of ESRF Beamline ID14-4, especially Joanna Timmins and Andrew McCarthy, for assistance with data collection and processing, and Monika Budayova-Spano of ILL for helping to make things possible.

#### References

- Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* **D52**, 30–42.
- Anslin, E. V. & Dougherty, D. A. (2006). *Modern Physical Organic Chemistry*. Sausalito: University Science Books.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.

- Chung, S.-T., Manis, J. J., McWethy, S. J., Patt, T. E., Witz, D. F., Wolf, H. J. & Wovcha, M. G. (1997). In *Biotechnology of Industrial Antibiotics*, edited by W. R. Strohl, pp. 165–186. New York: Dekker.
- Cowtan, K. (2006). *Acta Cryst.* **D62**, 1002–1011.
- Cowtan, K. (2008). *Acta Cryst.* **D64**, 83–89.
- Doublé, S. (1997). *Methods Enzymol.* **276**, 523–530.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Fauman, E. B., Blumenthal, R. M. & Cheng, X. (1999). *S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions*, edited by X. Cheng & R. M. Blumenthal, pp. 1–38. Singapore: World Scientific Publishing.
- French, S. & Wilson, K. (1978). *Acta Cryst.* **A34**, 517–525.
- Graaff, R. A. G. de, Hilge, M., van der Plas, J. L. & Abrahams, J. P. (2001). *Acta Cryst.* **D57**, 1857–1862.
- Halgren, T. A. (1996). *J. Comput. Chem.* **17**, 490–519.
- Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E. & Hutchison, G. R. (2012). *J. Cheminform.* **4**, 17.
- Hegazi, M. F., Borchard, R. T. & Schowen, R. L. (1976). *J. Am. Chem. Soc.* **98**, 3048–3049.
- Hegazi, M. F., Borchardt, R. T. & Schowen, R. L. (1979). *J. Am. Chem. Soc.* **101**, 4359–4365.
- Huang, Y., Komoto, J., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M. & Takusagawa, F. (2000). *J. Mol. Biol.* **298**, 149–162.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Koběrská, M., Kopecký, J., Olšovská, J., Jelínková, M., Ulanova, D., Man, P., Flieger, M. & Janata, J. (2008). *Folia Microbiol. (Praha)*, **53**, 395–401.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Li, A. J. & Nussinov, R. (1998). *Proteins*, **32**, 111–127.
- Liscombe, D. K., Louie, G. V. & Noel, J. P. (2012). *Nat. Prod. Rep.* **29**, 1238–1250.
- Magerlein, B. J. (1977). *Structure-Activity Relationships among the Semisynthetic Antibiotics*, edited by D. Perlman, pp. 600–651. New York: Academic Press.
- Martin, J. L. & McMillan, F. M. (2002). *Curr. Opin. Struct. Biol.* **12**, 783–793.
- McCarthy, A. A., Brockhauser, S., Nurizzo, D., Theveneau, P., Mairs, T., Spruce, D., Guijarro, M., Lesourd, M., Ravelli, R. B. G. & McSweeney, S. (2009). *J. Synchrotron Rad.* **16**, 803–812.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. (2009). *J. Comput. Chem.* **30**, 2785–2791.
- Murshudov, G. N., Dodson, E. J. & Vagin, A. A. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 93–104. Warrington: Daresbury Laboratory.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Najmanová, L., Kutejová, E., Kadlec, J., Polan, M., Olšovská, J., Benada, O., Novotná, J., Kameník, Z., Halada, P., Bauer, J. & Janata, J. (2013). *Chembiochem*, **14**, 2259–2262.
- Ness, S. R., de Graaff, R. A. G., Abrahams, J. P. & Pannu, N. S. (2004). *Structure*, **12**, 1753–1761.
- Olšovská, J., Jelínková, M., Man, P., Koběrská, M., Janata, J. & Flieger, M. (2007). *J. Chromatogr. A*, **1139**, 214–220.
- Pannu, N. S., McCoy, A. J. & Read, R. J. (2003). *Acta Cryst.* **D59**, 1801–1808.
- Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst.* **D54**, 1285–1294.
- Pannu, N. S. & Read, R. J. (2004). *Acta Cryst.* **D60**, 22–27.
- Pioszak, A. A., Murayama, K., Nakagawa, N., Ebihara, A., Kuramitsu, S., Shirouzu, M. & Yokoyama, S. (2005). *Acta Cryst.* **F61**, 867–874.
- Rajeswaran, M. & Srikrishnan, T. (2004). *Carbohydr. Res.* **339**, 2111–2115.
- Schubert, H. L., Blumenthal, R. M. & Cheng, X. (2003). *Trends Biochem. Sci.* **28**, 329–335.
- Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Cryst.* **D60**, 1355–1363.
- Skubák, P., Murshudov, G. N. & Pannu, N. S. (2004). *Acta Cryst.* **D60**, 2196–2201.
- Spížek, J. & Řezanka, T. (2004a). *Appl. Microbiol. Biotechnol.* **63**, 510–519.
- Spížek, J. & Řezanka, T. (2004b). *Appl. Microbiol. Biotechnol.* **64**, 455–464.
- Struck, A.-W., Thompson, M. L., Wong, L. S. & Micklefield, J. (2012). *Chembiochem*, **13**, 2642–2655.
- Trott, O. & Olson, A. J. (2010). *J. Comput. Chem.* **31**, 455–461.
- Ulanova, D., Novotná, J., Smutná, Y., Kameník, Z., Gažák, R., Šulc, M., Sedmera, P., Kadlčík, S., Plháčková, K. & Janata, J. (2010). *Antimicrob. Agents Chemother.* **54**, 927–930.
- Weissman, K. J. (2007). *Trends Biotechnol.* **25**, 139–142.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* **D57**, 122–133.
- Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
- Woodard, R. W., Tsai, M.-D., Floss, H. G., Crooks, P. A. & Coward, J. K. (1980). *J. Biol. Chem.* **255**, 9124–9127.
- Wright, J. L. (1983). *Biochemistry and Genetic Regulation of Commercially Important Antibiotics*, edited by L. C. Vining, pp. 311–328. London: Addison-Wesley.