Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural characterization of CalO1: a putative orsellinic acid methyltransferase in the calicheamicin-biosynthetic pathway

Aram Chang,[a,b] Shanteri Singh,[c] Craig A. Bingman,[b] Jon S. Thorson[c]* and George N. Phillips Jr[a,b]*

[a]Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706, USA, [b]Center for Eukaryotic Structural Genomics, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706, USA, and [c]Laboratory for Biosynthetic Chemistry, Pharmaceutical Sciences Division, School of Pharmacy, National Cooperative Drug Discovery Program, University of Wisconsin-Madison, 777 Highland Avenue, Madison, Wisconsin 53705, USA

Correspondence e-mail: jsthorson@pharmacy.wisc.edu phillips@biochem.wisc.edu

Received 26 November 2010
Accepted 28 January 2011

PDB Reference: CalO1, 3lst.

The X-ray structure determination at 2.4 Å resolution of the putative orsellinic acid C3 *O*-methyltransferase (CalO1) involved in calicheamicin biosynthesis is reported. Comparison of CalO1 with a homology model of the functionally related calicheamicin orsellinic acid C2 *O*-methyltransferase (CalO6) implicates several residues that are likely to contribute to the regiospecificity of alkylation. Consistent with the proposed requirement of an acyl-carrier-protein-bound substrate, this structural study also reveals structural determinants within CalO1 that are anticipated to accommodate an association with an acyl carrier protein.

1. Introduction

Given their unique structures, reactivities and therapeutic potential, the enediynes often stand out among natural products. There are currently two naturally occurring enediyne structural subfamilies: the chromoprotein (or nine-membered) enediynes and the ten-membered enediynes (Thorson *et al.*, 2000; Galm *et al.*, 2005). The chromoprotein enediynes share a bicyclo[7.3.0]enediyne nine-membered core and usually require an enediyne-sequestering/stabilizing protein, while the ten-membered enediynes are standalone small molecules that contain a bicyclo[7.3.1]enediyne core. All enediynes are highly efficient DNA/RNA-damaging agents and their corresponding producing organisms sometimes rely upon novel self-resistance mechanisms as exemplified by the 'self-sacrifice' calicheamicin (CLM) resistance protein CalC (Biggins *et al.*, 2003; Singh *et al.*, 2006). As illustrated by the ten-membered enediyne CLM γ_1^I (**5** in Fig. 1), DNA cleavage is facilitated *via* binding to the minor groove of the target DNA and results from DNA-backbone hydrogen abstraction by the highly reactive biradical intermediate of enediyne cycloaromatization (Zein *et al.*, 1988, 1989; Kumar *et al.*, 1997; Walker *et al.*, 1993). Notably, CLM also stands out as the 'cargo' of the first FDA-approved antibody–drug conjugate (the anti-CD33-CLM conjugate Mylotarg; Wu & Senter, 2005; Damle, 2004; Hamann, 2005) and, while Mylotarg has recently been removed from the market owing to limited efficacy and off-target toxicities (Thayer, 2010), the progression of alternative CLM-based conjugates continues to fuel an interest in enediyne synthetic and biosynthetic diversification (Zhang *et al.*, 2006; Kennedy *et al.*, 2007; Shen *et al.*, 2003; Jones & Fouad, 2002).

Early precursor labeling studies implicated distinct biosynthetic pathways for the nine- and ten-membered enediynes

(Lam *et al.*, 1993; Hensens *et al.*, 1989; Tokiwa *et al.*, 1992). However, the more recent cloning and characterization of several enediyne-encoding biosynthetic gene loci support a unified polyketide paradigm for enediyne-core biosynthesis in which the biosyntheses of both nine-membered and ten-membered enediynes diverge from a common biosynthetic intermediate (Ahlert *et al.*, 2002; Liu *et al.*, 2002, 2003, 2005; Zazopoulos *et al.*, 2003; Van Lanen *et al.*, 2007; Gao & Thorson, 2008; Zhang *et al.*, 2008; Horsman *et al.*, 2010). In addition to the polyketide core, most enediynes also contain additional polyketide-derived components (*e.g.* the orsellinic acid moiety of CLM). The biosynthesis of such precursors (**1** in Fig. 1) relies upon a unique set of iterative type I polyketide synthases (PKSs) that are reminiscent of the fungal 6-methylsalicyclic acid synthase (6-MSAS; Shen, 2003) and are subsequently modified *via* enzymatic oxidation, halogenation and/or methylation (**2**–**4** in Fig. 1). While bacterial iterative type I PKSs are now known to participate in the biosynthesis of at least five diverse natural products, avilamycin (Gaisser *et al.*, 1997), CLM (Ahlert *et al.*, 2002), neocarzinostatin (Liu *et al.*, 2005; Sthapit *et al.*, 2004), chlorothricin (Jia *et al.*, 2006; Shao *et al.*, 2006) and maduropeptin (Van Lanen *et al.*, 2007), the subsequent downstream modification of the resulting polyketide products remains poorly understood.

In the case of CLM, a total of six enzymes (CalO1–CalO6) are postulated to participate in orsellinic acid maturation. Of the postulated tailoring enzymes, the putative methyltransferase CalO6 has been demonstrated to efficiently methylate the C2 OH of *N*-acetyl cysteaminyl orsellinic acid (SNAc-orsellinic acid) *in vitro*, while the corresponding homolog CalO1 was incapable of turnover under identical conditions (S. Singh & J. S. Thorson, unpublished work). These preliminary studies, in conjunction with CalO2 ligand-binding studies and CalO2 structural studies which revealed a requirement for thioester-conjugated (*e.g.* acyl carrier protein) iodinated substrates (McCoy *et al.*, 2009), are consistent with the sequence of biosynthetic events highlighted in Fig. 1 and implicate CalO1 as the putative orsellinic acid C3 *O*-methyltransferase in this pathway.

As an extension of these studies, we report the X-ray structure determination of the putative CLM orsellinic acid C3 *O*-methyltransferase (CalO1) bound to cofactor (*S*-adenosylhomocysteine; SAH). These studies reveal insights regarding possible acyl carrier protein interaction and putative regiospecificity determinants of CalO1 and CalO6.

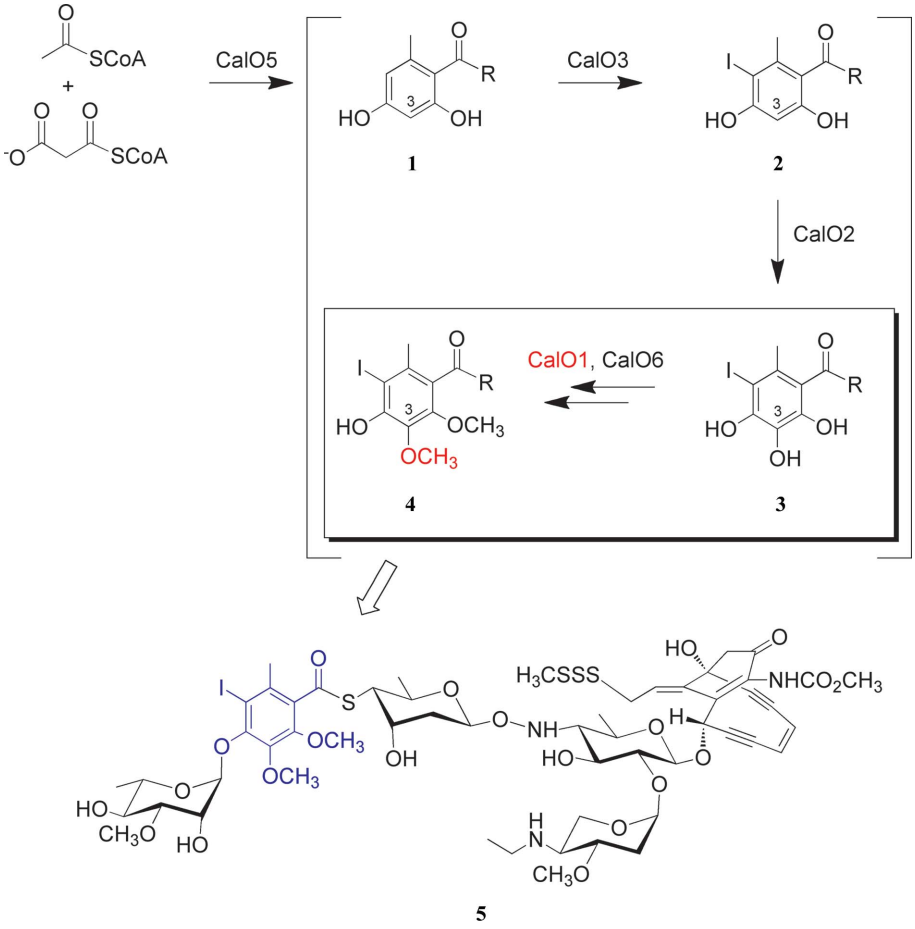

Figure 1
The proposed pathway for orsellinic acid biosynthesis and modification en route to calicheamicin (**5**) construction. The putative C3 OH methylation reaction catalyzed by CalO1 is highlighted within a box (red) and the mature orsellinate progeny in the context of calicheamicin is emphasized in blue. Based upon the ligand-binding studies described, *R* is anticipated to be an ACP or CoA thioester.

2. Materials and methods

2.1. Materials

Escherichia coli BL834 (DE3) competent cells were purchased from Invitrogen (Carlsbad, California, USA). The pET-28b *E. coli* expression vector and thrombin were purchased from Novagen (Madison, Wisconsin, USA). Primers were purchased from Integrated DNA Technology (Coralville, Iowa, USA). *Pfu* DNA polymerase was purchased from Stratagene (La Jolla, California, USA). Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs (Ipswich, Massachusetts, USA). Benzamidine Sepharose was purchased from Pharmacia. All other chemicals were of reagent grade or better and were purchased from Sigma (St Louis, Missouri, USA).

2.2. Gene cloning and expression, and purification of CalO1

The *calO1* gene from the calicheamicin producer, *Micromonospora echinospora* LL6600, was amplified from cosmid pJT1192c using the primer

pairs 5′-AGAGCAGTCATATGCAGCGGCAGCGTCCGC-CATCACGGGCCGG-3′ (forward, *Nde*I) and 5′-CGGTCA-TGTCCATCGCGGTCGGCGTGCCGGCCTGAGAATTC-3′ (reverse, *Eco*RI) and *Pfu* DNA polymerase. PCR products were digested with *Nde*I/*Eco*RI and ligated into the pET28a expression vector (*Nde*I/*Eco*RI; to generate the N-terminal His tag).

SeMet-labeled N-His_6-CalO1 was overproduced in *E. coli* B834 (DE3) using the Studier auto-induction method (Studier, 2005). A single transformant of *E. coli* B834 (DE3)/CalO1-pET28a was inoculated into 25 ml non-inducing medium supplemented with 35 mg ml^{-1} kanamycin and grown at 298 K overnight. The precultures were inoculated into 1 l auto-induction medium with 35 mg ml^{-1} kanamycin and were continuously grown for 24 h. The cells were collected by centrifugation (15 min, 5000*g*) and resuspended in buffer *A* (20 m*M* NaH_2PO_4, 300 m*M* NaCl, 10 m*M* imidazole pH 7.5). The cells were lysed *via* incubation with 1 mg ml^{-1} lysozyme (∼50 000 U mg^{-1}; Sigma, St Louis, Missouri, USA) for 30 min on ice followed by sonication (VirSonic 475; Virtis, Gardiner, New York, USA; 100 W, 4 × 30 s pulses, ∼1 min between pulses) on ice. N-His_6-CalO1 was purified *via* affinity chromatography (5 ml HiTrap Ni–NTA chelating column, GE Healthcare, Piscataway, New Jersey, USA) following a standard protocol with a linear imidazole (10–500 m*M*) elution gradient to provide 60 mg N-His_x-CalO1. The purified protein was desalted using a PD-10 column (GE Healthcare) with elution buffer (20 m*M* NaH_2PO_4, 50 m*M* NaCl pH 7.5). The N-His_6 tag was removed by thrombin cleavage following a standard protocol (Novagen) using 20 units of thrombin in 1× cleavage buffer for 20 h at room temperature per 60 mg N-His_6-CalO1. This was then treated with 1 ml Benzamidine Sepharose (Pharmacia) slurry to remove the thrombin. The His-tag-cleaved CalO1 was subsequently purified *via* a second round of affinity chromatography (5 ml HiTrap Ni–NTA chelating column; GE Healthcare, Piscataway, New Jersey, USA) and desalted using a PD-10 column (GE Healthcare) with elution buffer (20 m*M* NaH_2PO_4, 50 m*M* NaCl pH 7.5). The purified CalO1 was subsequently concentrated to 20 mg ml^{-1} with 5 m*M* SAH, flash-frozen in liquid nitrogen and stored at 193 K. Protein concentrations were determined by Bradford assay (Bio-Rad, Hercules, California, USA) using BSA as a standard.

2.3. Protein crystallization

Initial crystallization screening was performed with a local screen (UW192) as well as the commercial screens Index HT and SaltRX HT (Hampton Research, Aliso Viejo, California, USA) using a Mosquito dispenser (TTP LabTech, Royston, Herts, England) and the sitting-drop method. A local organics additive screen (CESG) was included as part of this initial screen. Crystal growth was monitored using Bruker Nonius Crystal Farms at 293 and 277 K. Diffraction-quality crystals of CalO1 were grown using the hanging-drop vapor-diffusion method. The reservoir solution consisted of 20% PEG MME 5K (polyethylene glycol monomethyl ether), 200 m*M* glycine, 100 m*M* bis-tris propane (BTP) pH 7.0 and 1 m*M* CHAPS {3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate}. The hanging drop consisted of 1 µl protein solution (20 mg ml^{-1} containing 5 m*M* SAH) mixed with 1 µl reservoir solution. Crystallization trays were stored at 293 K. Under these conditions, CalO1 crystals required one month to reach full size (100 × 15 × 15 µm). The crystals were subsequently soaked in increasing concentrations of ethylene glycol in mother liquor up to a final concentration of 20%(*v*/*v*) and flash-frozen in a stream of liquid nitrogen.

Table 1
Crystal parameters and data-collection, phasing and refinement statistics.

Values in parentheses are for the highest resolution shell.

Crystal parameters		
Space group	$C222_1$	
Unit-cell parameters (Å, °)	a = 63.515, b = 93.603, c = 240.971, $\alpha = \beta = \gamma$ = 90.00	
Data-collection statistics		
Wavelength (Å)	0.9794	0.9641
Resolution range (Å)	50.00–2.40	50.00–2.40
No. of reflections (measured/unique)	52781/28057	51604/27589
Completeness (%)	97.7 (83.4)	95.9 (74.8)
R_{merge}†	0.120 (0.525)	0.116 (0.721)
Multiplicity	12.3 (9.0)	11.7 (6.7)
Mean $I/\sigma(I)$	22.53 (3.57)	21.3 (2.2)
Phasing statistics‡		
Mean FOM (centric/acentric)	0.305/0.462	
Phasing power ANO	1.696	
Phasing power ISO (centric/acentric)	0.794/0.894	
Cullis R factor (anomalous/isomorphous)	0.865/0.873	
Refinement and model statistics		
Resolution range (Å)	48.17–2.40	
No. of reflections (work/test)	27959/1411	
R_{cryst}§	0.203	
R_{free}¶	0.251	
R.m.s.d. bonds (Å)	0.004	
R.m.s.d. angles (°)	0.717	
B factor (protein/solvent) (Å^2)	51.03/41.58	
No. of protein atoms	5303	
No. of waters	181	
Ramachandran plot (%)		
Most favorable region	98.79	
Additional allowed region	1.21	
Generously allowed region	0.0	
Disallowed region	0.0	
PDB code	3lst	

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual measurement of the reflection and $\langle I(hkl)\rangle$ is the mean intensity of this reflection. ‡ Phasing in *autoSHARP*. § $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively. ¶ R_{free} was calculated as R_{cryst} using approximately 5% of randomly selected unique reflections that were omitted from the structure refinement.

2.4. Diffraction data collection and structure determination

X-ray diffraction data were collected on the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) beamlines at the Advanced Photon Source at Argonne National Laboratory (Argonne, Illinois, USA). Data sets were collected using X-ray wavelengths of 0.9794 and 0.9642 Å. The data sets were indexed and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). *phenix.hyss* and *SHELXD* were utilized to determine the selenium substructure (Adams *et al.*, 2010; Sheldrick, 2008), *autoSHARP* was utilized for

MAD phasing, *DM* was utilized for density modification and *ARP*/*wARP* was utilized for automatic model building (de La Fortelle & Bricogne, 1997; Perrakis *et al.*, 1999; Cowtan & Main, 1996). The structure was completed by alternating rounds of manual model building with *Coot* and refinement with *phenix.refine* (Emsley & Cowtan, 2004; Adams *et al.*, 2010). TLS groups, which were selected based on the output of the *TLSMD* web server, were incorporated during the final stage of refinement (Painter & Merritt, 2006*a*,*b*). Structure quality was assessed using *PROCHECK* and *MolProbity* (Laskowski *et al.*, 1993; Chen *et al.*, 2010). Crystallographic statistics are summarized in Table 1. All figures were generated by *PyMOL* (DeLano, 2002).

3. Results and discussion

3.1. Structure quality and overall fold

The crystal structure of CalO1 with bound SAH was solved at 2.4 Å resolution (the data statistics are listed in Table 1). There are two CalO1 molecules in the asymmetric unit and both chains have well defined electron density for residues 13–345 (Fig. 2*a*) The r.m.s.d. of C^{α}-atom positions between superposed chains *A* and *B* is 1.1 Å. The primary differences are located in the C-terminal domain, where the *A* chain shows a more closed conformation compared with the *B* chain. The *PISA* server (Krissinel & Henrick, 2007) predicts that each subunit in the asymmetric unit represents a half dimer, with the biologically relevant dimer generated by a crystallographic twofold axis. The total buried surface area of each dimer is 8500 $Å^2$, which is approximately 30% of the total surface area. The calculated electrostatic surface rendering illustrates that dimerization is mediated through electrostatic interactions between the N-terminal domains (DeLano, 2002; Fig. 2*b*).

3.2. Comparison with other methyltransferases

CalO1 is comprised of two domains. The N-terminal domain (Met1–Ala135), which is composed of α-helices, is involved in formation of the dimer. The C-terminal domain (Phe136–Ala345), which contains the solvent-exposed active site, adopts an α/β Rossmann fold and is predominately involved in cofactor/substrate binding. The *DALI* server (Holm *et al.*, 2008) calculates that caffeic acid methyltransferase (ChOMT; *Z* score 32.9, r.m.s.d. 2.5 Å; Zubieta *et al.*, 2001), mitomycin methyltransferase (MmcR; *Z* score 32.1, r.m.s.d. 3.7 Å; PDB entry 3gxo; S. Singh, A. Chang, C. A. Bingman, G. N. Phillips

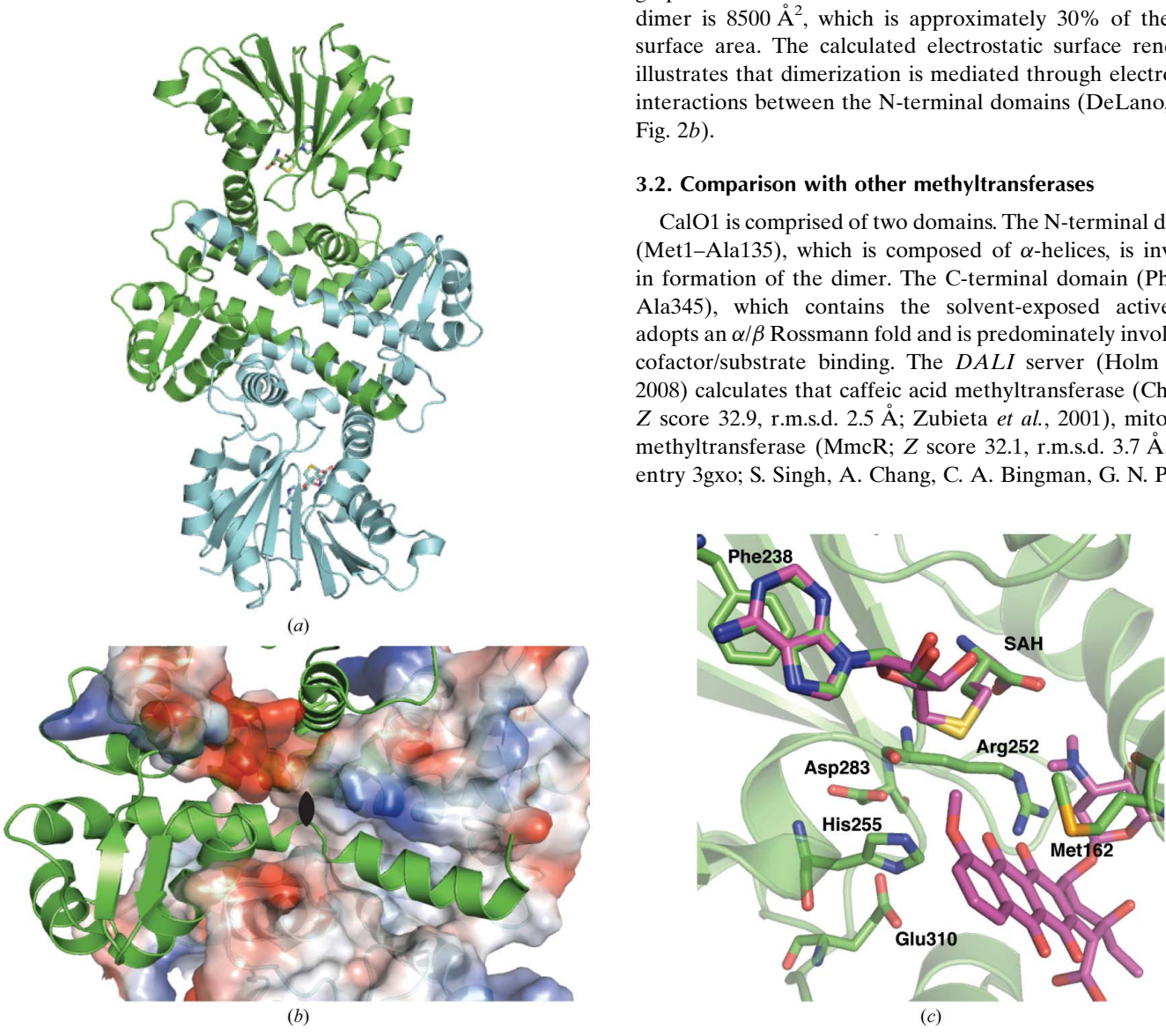

Figure 2
(*a*) The CalO1 dimer. In this representation bound SAH is represented as a stick model, one half of the dimer observed in the asymmetric unit is highlighted in green and the cyan-colored chain represents that derived from the symmetry operator $x, -y+1, -z+1$. (*b*) The dimer interface of CalO1. In this representation the electrostatic surface is illustrated for one subunit, the color indicates the potential charge (negative, red; positive, blue) and the twofold axis is indicated by the crystallographic symbol. (*c*) The SAH- and substrate-binding region of CalO1. In this model the C-terminal domains of CalO1 and DnrK (PDB entry 1tw2) were aligned to identify the probable substrate-binding site of CalO1. CalO1-bound SAH and corresponding residues are colored green, while DnrK-bound SAH, substrate (4-methoxy-ε-rhodomycin T) and corresponding residues are colored magenta.

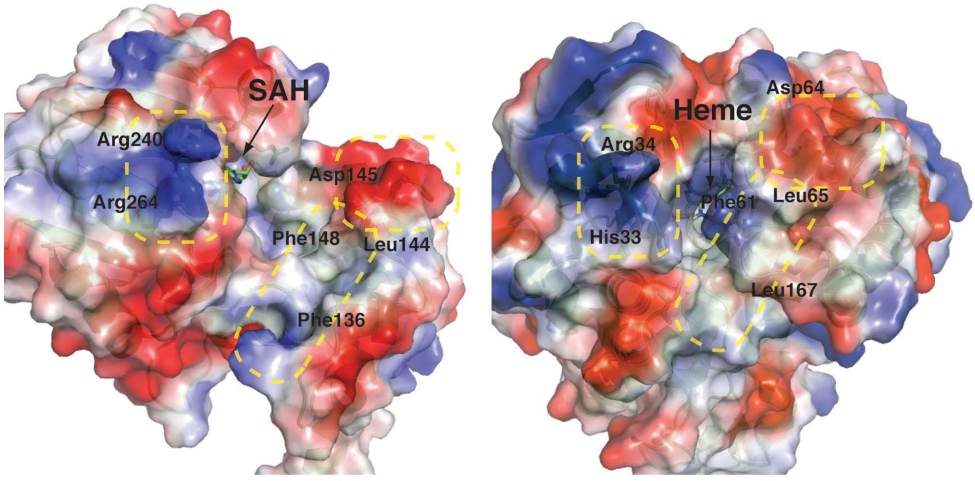

Figure 3
(*a*) Electrostatic surface diagram around the CalO1 active site. Putative binding regions are marked as yellow dotted lines. Within the proposed ACP-binding region, Arg240 and Arg264 comprise the positively charged region and Phe136, Leu144 and Phe148 comprise the hydrophobic region. (*b*) Electrostatic surface diagram around the CalO2 active site (PDB entry 3buj; McCoy *et al.*, 2009). His33 and Arg34 comprise the positively charged region and Phe61, Leu65 and Leu167 comprise the hydrophobic region.

Jr & J. S. Thorson, unpublished work) and isoliquiritigenin methyltransferase (IOMT; *Z* score 32.0, r.m.s.d. 2.3 Å; Zubieta *et al.*, 2001) are structurally similar to CalO1. Consistent with this, all three of the structural homologs are small-molecule O-methyltransferases from plants (ChoMT and IOMT) or bacteria (MmcR) which alkylate aromatic hydroxyl groups.

3.3. Cofactor and substrate binding

Well defined electron density was observed for a CalO1-bound SAH molecule in the C-terminal domain adjacent to the conserved DVGGG*X*G motif (Fig. 2*c*, Supplementary Fig. 1[1]; Jansson *et al.*, 2004). Based on this structure, the key CalO1–cofactor interactions include (i) Phe238–SAH adenine π-stacking, (ii) an Asp212–SAH adenine N6 hydrogen bond, (iii) an Arg213–SAH ribose O2 hydrogen bond and (iv) a Lys251–SAH carboxyl hydrogen bond. Most of these residues and interactions are highly conserved in other natural product methyltransferase structures (Zubieta *et al.*, 2001; Jansson *et al.*, 2004).

Likewise, the substrate-binding site and the reaction mechanism are highly conserved among natural product O-methyltransferases. Based on these similarities, we propose His255 to be the CalO1 catalytic residue, with Asp283 and Glu310 mediating the proper orientation of His255 for deprotonation of the substrate C3 hydroxyl group. Alignments with the substrate-bound structures of DnrK or MmcR reveal Met162, Arg252, Phe113, Met298, Met301 and Met302 as potential contributors to substrate recognition/orientation (Fig. 2*c*). Consistent with the latter three residues in this series, similarly organized methionines have been noted to participate in sequestering aromatic substrates in the active site of methyltransferases, where they contribute to the hydrophobicity of the active site and possibly to stabilize the approach of pantetheinate conjugates (Zubieta *et al.*, 2001).

While the putative substrate-binding cavity is hydrophobic and solvent-exposed, the distance from His255 to the CalO1 outer surface is approximately 18 Å, implicating the participation of a pantetheinate arm in the CalO1 reaction (Fig. 3*a*). The CalO2–CalO5 ACP simulated docked model proposed electrostatic interaction *via* positive (His33 and Arg34) and neutral (Phe61 and Leu65) regions of CalO2 (Fig. 3*b*; McCoy *et al.*, 2009). Consistent with the general structure of an ACP (a three- or four-helix bundle with acidic and hydrophobic regions that interact with complementary basic and hydrophobic regions of interacting proteins; Lai *et al.*, 2006), the CalO1 electrostatic surface rendering exhibits similar positive (Arg240 and Arg264) and neutral (Phe136, Leu144 and Phe148) properties near the active site (Fig. 3*a*).

3.4. Comparison with CalO6 and implications for methyltransferase regiospecificity

CalO1 and CalO6 share 31% sequence identity and, as expected, the CalO1 and CalO6 active sites and key residues that are involved in substrate/cofactor recognition and catalysis are highly conserved. Exceptions to this high level of conservation between the active sites of CalO1 and CalO6 include the substitution of Phe113, Arg252, Met298 and Met301 of CalO1 in CalO6 (by Leu103, Ser249, Asn299 and Asn302, respectively). Considering that the CalO1 and CalO6 substrates share a similar molecular structure, these residue differences are proposed to affect substrate reorientation and thereby to dictate the differences in regiospecificity between the two enzymes. A homology model of CalO6 was generated by *MolWeb* (Eswar *et al.*, 2007) and illustrates the key similarities and differences between the CalO1 and CalO6 active sites (Fig. 4).

4. Conclusions

We have determined the crystal structure of CalO1, a putative orsellinic acid methyltransferase from the calicheamicin-biosynthetic pathway, which exhibits a common fold for small-molecule O-methyltransferases. Structural analysis of CalO1 indicates that the reaction might require ACP-bound substrate. Comparison with the CalO6 homology model, another

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: YT5031). Services for accessing this material are described at the back of the journal.

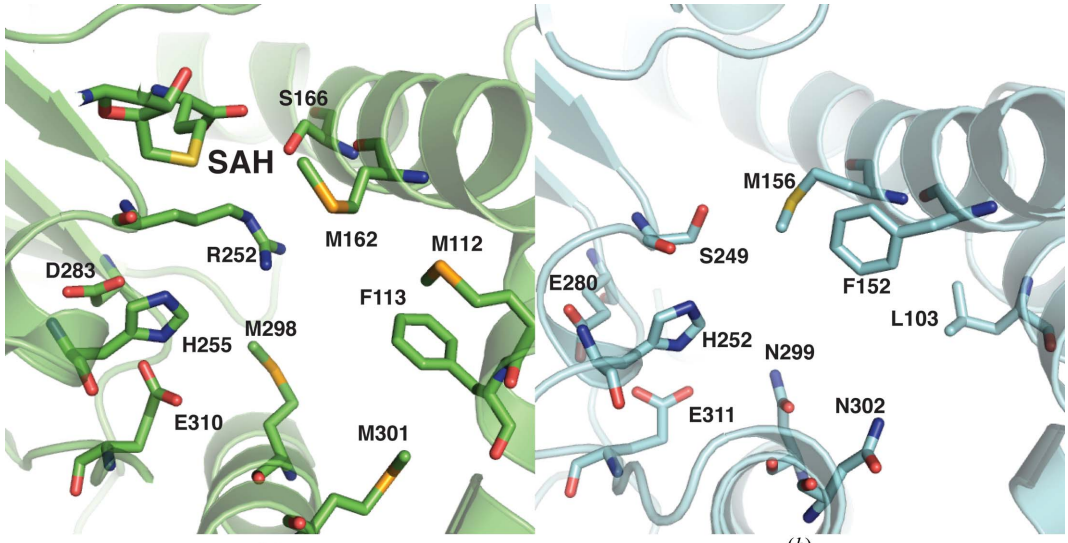

Figure 4
Comparison of the substrate-binding regions of CalO1 (*a*) and the CalO6 homology model (*b*). The CalO6 homology model was generated with *MolWeb* (Eswar *et al.*, 2007) using the sequence of CalO6. Proposed catalytic residues are conserved between CalO1 and CalO6 (His255, Asp283 and Glu310 in CalO1; His252, Glu280 and Glu311 in CalO6), while other residues proposed to position the substrates show differences (Phe113, Arg252, Met298 and Met301 in CalO1; Leu103, Ser249, Asn299 and Asn302 in CalO6).

putative orsellinic acid methyltransferase from the same pathway, suggests several residues that contribute to the regiospecificity of the two enzymes.

The authors thank Christopher M. Bianchetti and Dr E. Sethe Burgie for helpful discussions. This research was supported in part by National Institutes of Health Grant CA84374 (JST) and U54 GM074901 (NIGMS) and National Institutes of Health Molecular Biophysics Training grant GM08293 (AC). JST is a UW HI Romnes Fellow and holds the Laura and Edward Kremers Chair in Natural Products. The General Medicine and Cancer Institute Collaborative Access Team (GM/CA-CAT) has been funded in whole or in part with federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Science (Y1-GM-1104). Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science under contract W-31-102-ENG-38.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B. O., Huang, K., Fonstein, L., Czisny, A., Whitwam, R. E., Farnet, C. M. & Thorson, J. S. (2002). *Science*, **297**, 1173–1176.
Biggins, J. B., Onwueme, K. C. & Thorson, J. S. (2003). *Science*, **301**, 1537–1541.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.
Damle, N. K. (2004). *Expert Opin. Biol. Ther.* **4**, 1445–1452.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. & Sali, A. (2007). *Curr. Protoc. Protein Sci.* **50**, 2.9.1–2.9.31.
Gaisser, S., Trefzer, A., Stockert, S., Kirschning, A. & Bechthold, A. (1997). *J. Bacteriol.* **179**, 6271–6278.
Galm, U., Hager, M. H., Van Lanen, S. G., Ju, J., Thorson, J. S. & Shen, B. (2005). *Chem. Rev.* **105**, 739–758.
Gao, Q. & Thorson, J. S. (2008). *FEMS Microbiol. Lett.* **282**, 105–114.
Hamann, P. R. (2005). *Expert Opin. Ther. Pat.* **15**, 1087–1103.
Hensens, O. D., Giner, J. L. & Goldberg, I. H. (1989). *J. Am. Chem. Soc.* **111**, 3295–3299.
Holm, L., Kääriäinen, S., Rosenström, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.
Horsman, G. P., Chen, Y., Thorson, J. S. & Shen, B. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 11331–11335.
Jansson, A., Koskiniemi, H., Mäntsälä, P., Niemi, J. & Schneider, G. (2004). *J. Biol. Chem.* **279**, 41149–41156.
Jia, X.-Y., Tian, Z.-H., Shao, L., Qu, X.-D., Zhao, Q.-F., Tang, J., Tang, G.-L. & Liu, W. (2006). *Chem. Biol.* **13**, 575–585.
Jones, G. B. & Fouad, F. S. (2002). *Curr. Pharm. Des.* **8**, 2415–2440.
Kennedy, D. R., Gawron, L. S., Ju, J., Liu, W., Shen, B. & Beerman, T. A. (2007). *Cancer Res.* **67**, 773–781.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Kumar, R. A., Ikemoto, N. & Patel, D. J. (1997). *J. Mol. Biol.* **265**, 187–201.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lai, J. R., Koglin, A. & Walsh, C. T. (2006). *Biochemistry*, **45**, 14869–14879.
Lam, K. S., Veitch, J. A., Golik, J., Krishnan, B., Klohr, S. E., Volk, K. J., Forenza, S. & Doyle, T. W. (1993). *J. Am. Chem. Soc.* **115**, 12340–12345.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Liu, W., Ahlert, J., Gao, Q., Wendt-Pienkowski, E., Shen, B. & Thorson, J. S. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 11959–11963.
Liu, W., Christenson, S. D., Standage, S. & Shen, B. (2002). *Science*, **297**, 1170–1173.
Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S. D., Bae, J., Van Lanen, S. G., Zazopoulos, E., Farnet, C. M., Yang, C. F. & Shen, B. (2005). *Chem. Biol.* **12**, 293–302.
McCoy, J. G., Johnson, H. D., Singh, S., Bingman, C. A., Lei, I.-K., Thorson, J. S. & Phillips, G. N. (2009). *Proteins*, **74**, 50–60.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Painter, J. & Merritt, E. A. (2006*a*). *Acta Cryst.* D**62**, 439–450.
Painter, J. & Merritt, E. A. (2006*b*). *J. Appl. Cryst.* **39**, 109–111.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Shao, L., Qu, X.-D., Jia, X.-Y., Zhao, Q.-F., Tian, Z.-H., Wang, M., Tang, G.-L. & Liu, W. (2006). *Biochem. Biophys. Res. Commun.* **345**, 133–139.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Shen, B. (2003). *Curr. Opin. Chem. Biol.* **7**, 285–295.
Shen, B., Liu, W. & Nonaka, K. (2003). *Curr. Med. Chem.* **10**, 2317–2325.
Singh, S., Hager, M. H., Zhang, C., Griffith, B. R., Lee, M. S., Hallenga, K., Markley, J. L. & Thorson, J. S. (2006). *ACS Chem. Biol.* **1**, 451–460.
Sthapit, B., Oh, T.-J., Lamichhane, R., Liou, K., Lee, H. C., Kim, C.-G. & Sohng, J. K. (2004). *FEBS Lett.* **566**, 201–206.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Thayer, A. (2010). *Chem. Eng. News*, **88**, 12.
Thorson, J. S., Sievers, E. L., Ahlert, J., Shepard, E., Whitwam, R. E., Onwueme, K. C. & Ruppen, M. (2000). *Curr. Pharm. Des.* **6**, 1841–1879.
Tokiwa, Y., Miyoshisaitoh, M., Kobayashi, H., Sunaga, R., Konishi, M., Oki, T. & Iwasaki, S. (1992). *J. Am. Chem. Soc.* **114**, 4107–4110.
Van Lanen, S. G., Oh, T.-J., Liu, W., Wendt-Pienkowski, E. & Shen, B. (2007). *J. Am. Chem. Soc.* **129**, 13082–13094.
Walker, S., Murnick, J. & Kahne, D. (1993). *J. Am. Chem. Soc.* **115**, 7954–7961.
Wu, A. M. & Senter, P. D. (2005). *Nature Biotechnol.* **23**, 1137–1146.
Zazopoulos, E., Huang, K., Staffa, A., Liu, W., Bachmann, B. O., Nonaka, K., Ahlert, J., Thorson, J. S., Shen, B. & Farnet, C. M. (2003). *Nature Biotechnol.* **21**, 187–190.
Zein, N., Poncin, M., Nilakantan, R. & Ellestad, G. A. (1989). *Science*, **244**, 697–699.
Zein, N., Sinha, A. M., McGahren, W. J. & Ellestad, G. A. (1988). *Science*, **240**, 1198–1201.
Zhang, C., Griffith, B. R., Fu, Q., Albermann, C., Fu, X., Lee, I.-K., Li, L. & Thorson, J. S. (2006). *Science*, **313**, 1291–1294.
Zhang, J., Van Lanen, S. G., Ju, J., Liu, W., Dorrestein, P. C., Li, W., Kelleher, N. L. & Shen, B. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 1460–1465.
Zubieta, C., He, X.-Z., Dixon, R. A. & Noel, J. P. (2001). *Nature Struct. Biol.* **8**, 271–279.