

Ryan Latimer,^a Kateryna Podzelinska,^b Alexei Soares,^c Anupam Bhattacharya,^a Leo C. Vining,^d Zongchao Jia^b and David L. Zechel^{a*}

^aDepartment of Chemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada, ^bDepartment of Biochemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada, ^cDepartment of Biology, Brookhaven National Laboratory, Upton, USA, and ^dDepartment of Biology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

Correspondence e-mail:
 dlzechel@chem.queensu.ca

Received 21 November 2008
 Accepted 17 December 2008

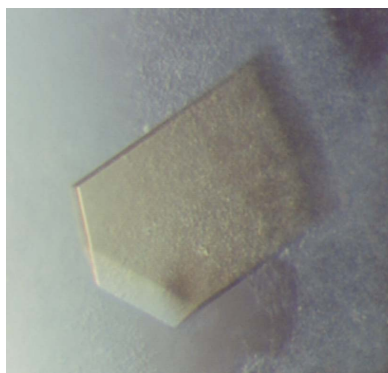
Expression, purification and preliminary diffraction studies of CmlS

CmlS, a flavin-dependent halogenase (FDH) present in the chloramphenicol-biosynthetic pathway in *Streptomyces venezuelae*, directs the dichlorination of an acetyl group. The reaction mechanism of CmlS is of considerable interest as it will help to explain how the FDH family can halogenate a wide range of substrates through a common mechanism. The protein has been recombinantly expressed in *Escherichia coli* and purified to homogeneity. The hanging-drop vapour-diffusion method was used to produce crystals that were suitable for X-ray diffraction. Data were collected to 2.0 Å resolution. The crystal belonged to space group C2, with unit-cell parameters $a = 208.1$, $b = 57.7$, $c = 59.9$ Å, $\beta = 97.5^\circ$.

1. Introduction

Naturally produced organohalogenes often display potent bioactivities and accordingly serve as a rich source of new drugs (Gribble, 2003). The first enzymes shown to regiospecifically catalyze halogenation were the flavin-dependent halogenases (FDHs; Keller *et al.*, 2000). Since this seminal discovery, a number of other enzymes that catalyze regiospecific and stereospecific halogenation have been discovered (Blasiak & Drennan, 2008; Neumann *et al.*, 2008). The FDHs have received particular attention since they are capable of halogenating a diverse array of natural products. The structural characterization of FDHs is still in its infancy, with only four structures known to date: PrnA (Dong *et al.*, 2005), RebH (Yeh *et al.*, 2007), *Shewanella frigidimarina* halogenase (PDB code 2pyx) and CndH (Buedenbender *et al.*, 2008). The first three of these enzymes chlorinate tryptophan, yielding 7-chlorotryptophan, whilst the recently characterized CndH chlorinates the *ortho* position of a phenol ring during the biosynthesis of chondrochloren. Mechanistic studies have shown that the flavin cofactor of FDHs generates HOCl, which is believed to either form a stable chloroamine intermediate (Yeh *et al.*, 2007) or hydrogen bond (Flecks *et al.*, 2008) to a conserved Lys residue in the active site (Fig. 1). The Lys residue in turn directs regiospecific chlorination of the substrate indole ring through an electrophilic aromatic substitution (EAS) reaction. Interestingly, the residues that are proposed to stabilize the carbocation intermediate (Dong *et al.*, 2005) are not conserved in FDH homologues (Fig. 1), although many of these enzymes also catalyze EAS reactions (Buedenbender *et al.*, 2008). Therefore, a crucial question is how the FDH family adopts a conserved halogenation machinery to react with a remarkable array of substrates such as indole (Dong *et al.*, 2005; Yeh *et al.*, 2007), pyrrole (Kirner *et al.*, 1998), quinone (Winter *et al.*, 2007), phenyl (Buedenbender *et al.*, 2008) and alkynyl (Edwards *et al.*, 2004) groups.

A unique addition to this list of functional group conversions is provided by CmlS, a FDH that is present in the chloramphenicol-biosynthetic pathway found in *Streptomyces venezuelae* (Pirae *et al.*, 2004). CmlS appears to catalyze what resembles a classical haloform reaction on an acetyl group, with the exception that the reaction stops after two halogenation events, producing the dichloroacetyl moiety on chloramphenicol. A structural view of the CmlS active site is



	FAD	
CmlS	---MTRSKVAIIGGGPAGSVAGLTLHK---LGHDVTIYERSAFPYRVGESLIPGTMISI	53
CrpH (27%)	MSTLPNSTQILIIIGGGPSGSTAATLLAR---EGFDVTLLEREVFPRVHVGESLIPSALEI	57
PrnA (11%)	--MNKPKIKNIVIVGGGTAGWMAASYLVRALQQQVNIITLIESAAIPRIGVGEATIPSLQKV	58
RebH (13%)	--MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATIPNLQTA	58
	: : * : * * * : * * : : : : * : * * * : * :	
CmlS	LN-RLGLQEK--IDAQNYVKIPSAATFLWG-----QDQAPWTFSAFAAPKV	94
CrpH	FD-LLGVREK--IEAYGFQRKPGAYIEWG-----TEKWSLNFGEITGDN	98
PrnA	FFDFLGIPEREWMPQVNGAFKAAIKFVNWRKSPD-----PSREDFYFHLFGSVPNCD	110
RebH	FFDFLGIPEDEWMRECNASYKVAIKFINWRTAGEGTSEARELDGGPDHFYHSFGLLKYHE	118
	: * : * : : . * : : : : :	
CmlS	APWVFDHAVQVKREE-----FDKLLL	115
CrpH	T-----YSFQVRRDE-----FDHLLL	114
PrnA	GVPLTHYWLKRKREQGF-QQPMAYACYPPGALDGKLPCLADGTRQMS-HAWHFD AHLVA	168
RebH	QIPLSHYWFDRSYRGKTVEPFYACYKEPVIL DANRSRRLDGSKVTN-YAWHFD AHLVA	177
	: . : * :	
CmlS	DEARSRG-ITVHEETPVTDV DLS-DPDRVVLTVRRGGESVTVESDFVIDAGG-SGGPISR	172
CrpH	EHSK SQG-KVVFEGTKIRQLSPDGRPRSATWSQSNDDTGEISFDFMIDASGRAGIMATE	173
PrnA	DFLKRWA-VERGVNRVDEVEVQ LNDRGYISTLLTKEGRTLEADLFIDCSGMRGLLINQ	227
RebH	DFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTATGRVFDADLFDVDCSGFRGLL INK	237
	: : . : : : : * : : * * * * * * * :	
	Isoalloxazine	
CmlS	KLGV--RQYDEFYRNFAVWSYFKLKDPPFEGDLKGTYSITFEDGWVWMIPIKDDLYSVGL	230
CrpH	YLKN--RRLHDVFQNVGIWGYWKNALRLPKGQSGAIALGSI PDGWVWGI PLDEEIMSVGV	231
PrnA	ALKEPFIDMSDYLLCDSAVASAVPNDDAREGVEPYTSAIAMNSGWTWKIPMLGRFGSGYV	287
RebH	AMEEPFLDMSDHLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFGTGYV	297
	: : : . . . : : : * * * * * * * : : :	
CmlS	VVDRSKSAEVREQGADAFYSSTLAKCAKAMDILGGAEQVDEVRIQDWSYDTEVFSADRF	290
CrpH	VMHKSTYKERLTKNLKDIYVEAIAECP LIADLVALGELVSDVKVEQDYSYTSDFSFGPAY	291
PrnA	FSSKFTSRDQATADFLKLWG-----LSDNQQLNQIKFRVGRNKRRAWVNNCVSI	335
RebH	YSSRFATEDEAVREFCEMWHLD-----P-ETQPLNRIRFRVGRNRRRAWVGNVCVSI	346
	: : : : : : : : : : : :	
	↓ ↓ Cl ⁻	
CmlS	FLCGDAACFTDPLFSQGVHLASQSAVSAASAAAIDRITR HGDEKDAVHAWYNRTRYEAYEQY	350
CrpH	FISGDAACFLDPLSSGVHLATYSALLAAASITSVIRGEVTEQAASFYDQSYRQAYLRF	351
PrnA	GLS--S-CFLEPLESTGIYFYIYAALYQLVKHFPDTS----FDPRLRDAFNAEIVYMFDDC	388
RebH	GTS--S-CFVPEPLESTGIYFVYAALYQLVKHFPDKS----LNPVLTARFNREIETMFD D T	399
	. * * : * * * * : : : : : : : : : :	
CmlS	HQFLASFYTFASFTEPDSEFWKRKRITESDDDRLTRKKWFESLAGNGPEDPSGTVASFRD	410
CrpH	LVFVSAFYDQN--RGKDSYFWEAQRLSRRDFGSSNLKLAFLNLVSG-----395	
PrnA	RDFVQAHFYFTT--SREDTPFWLANRHELRLSDAIQEKVERYKAGLP-----432	
RebH	RDFIQAHFYFS--PRDTPFWRAN-KELRLADGMQEKIDMYRAGMAIN-----444	
	* : : : : * : * * . . *	
CmlS	RASTMIAIGRHQRPELSDDFSEAE LNPARVRWISDLTKRLNSITRFKWTGGKAVLKQHYR	470
CrpH	-----VEDLEDAKEGIADFVMAEMSQR IQSSH S	423
PrnA	-----LTTTSFDDSTYYETFDYEFKNFWLNGNYICYIFAGLG	468
RebH	-----APASDDAQLYGNFEEEFNFWNNSNYCVLAGLG	479
	: * :	
CmlS	VEPIGFRLEQREVLANGEGLDMAQYPMDD EARQIFQDLAEEEFYKTLVKRLGAVGRQEL	530
CrpH	IRQDKQALAIEREKGNEMKTN AQFFNAVEGFSILSAVAVDGLYVTTQPKLGLVQVIPL	483
PrnA	MLPDRSLPLLQHRP-ESIQKAEAMFASIRREAERLRTSLPTNYDYLRSLRDGAQLSRNQH	527
RebH	LVPDAPSPRLAHMP-QATESVDEVFVGA VKDRQRNLL ETLP SLHEFLR-----QQH	528
	: . : : : : : :	
CmlS	STQIVVRLMEAGLLTG YDAQGEKVFVQGRLLHFGGVGV EYEV-----571	
CrpH	QRNSLLHT-----491	
PrnA	GPTLAAQERQ-----537	
RebH	GR-----530	

Figure 1

Alignment of the sequences of CmlS (as cloned in this work) and the flavin-dependent halogenases CrpH (GenBank accession No. ABM21576), PrnA (PDB code 2apg) and RebH (PDB code 2oa1). The alignment was performed with *ClustalW2*. Sequence percentage identities relative to CmlS are given in parentheses. Conserved regions of sequence are highlighted, including the FAD diphosphate-binding site (grey), the residues lining the FAD isoalloxazine-ring binding site and the tunnel that guides HOCl (yellow), the chloride-binding site (green) and the Lys residue which directs chlorination (red). The residues conserved in tryptophan halogenases PrnA and RebH which are thought to stabilize the carbocation intermediate of an electrophilic aromatic substitution (EAS) reaction are shown in turquoise. Note that CrpH also catalyzes an EAS reaction. The sequence IFRRSV in the CmlS sequence recorded in GenBank (accession No. AAK08979) is located between amino acids Pro302 and Leu303, which are highlighted with arrows (↓). The very high sequence conservation in this region suggests that the IFRRSV sequence in the CmlS GenBank entry is incorrect.

critical to determine the reaction mechanism. To this end, we report the expression and purification of CmlS and the generation of crystals that currently diffract to 2.0 Å resolution.

2. Materials and methods

2.1. Cloning, expression and purification

The gene encoding CmlS was PCR-amplified from the plasmid pJV526 (Pirae *et al.*, 2004) with the primers 5'-GCAGCCATATG-ACACGATCGAAGGTGGCGA-3' and 5'-CCGCAAGCTTTCAGACCTCGTACTCGAC-3' (*Nde*I and *Hind*III sites, respectively, are in bold). The purified PCR product was digested with *Nde*I and *Hind*III and ligated into similarly digested pET-28a (Novagen). The resulting plasmid, pET-28-CmlS, encodes CmlS with an N-terminal hexahistidine tag. Sequencing of both strands of pET-28-CmlS revealed that our *cmlS* clone differed from the *cmlS* sequence deposited in GenBank (accession No. AAK08979). The *cmlS* gene in pET-28-CmlS had two silent mutations (bases 702 and 948) and the DNA sequence from 907–924, which encodes the amino-acid sequence IFRRSV (residues 303–308 of AAK08979), was absent. A sequence alignment performed with our cloned CmlS amino-acid sequence revealed that the IFRRSV insertion would disrupt a highly conserved region shared by the FDH family (Fig. 1). This suggests that the IFRRSV sequence in GenBank accession No. AAK08979 is incorrect.

To express CmlS, *Escherichia coli* BL21(DE3) cells (Novagen) were transformed with pET-28-CmlS and grown at 310 K on solid Luria–Bertani medium containing 1% agarose and 50 µg ml⁻¹ kanamycin. A single colony was used to inoculate 5 ml Luria–Bertani medium supplemented with 50 µg ml⁻¹ kanamycin, which was then incubated overnight in an air shaker (225 rev min⁻¹, 310 K) to obtain a saturated culture. The saturated culture (5 ml) was used to inoculate 500 ml Luria–Bertani medium containing 50 µg ml⁻¹ kanamycin, which was incubated in an air shaker (225 rev min⁻¹, 310 K) until the optical density (OD₆₀₀) of the culture reached ~0.6. The cells were then induced with 0.5 mM IPTG and allowed to grow for 24 h at 288 K, after which time the cells were collected by centrifugation at 4000 rev min⁻¹ for 20 min at 277 K. The resulting cell pellet was stored at 253 K until purification. The cells were resuspended in 50 mM Tris–HCl, 2 mM DTT, 300 mM NaCl, 10 mM imidazole pH 7.5 and lysed using two passes through an EmulsiFlex C5 cell homogenizer (Avestin) at 138 MPa. The lysed cells were then centrifuged at 40 000g for 30 min at 277 K. The supernatant was loaded onto a nickel–nitrilotriacetic acid (Ni–NTA) column (Qiagen). CmlS was then eluted using an imidazole gradient ranging from 10 to 500 mM over ten column volumes at a flow rate of 5 ml min⁻¹ using an ÄKTA FPLC system (GE Healthcare). Fractions containing pure CmlS, as shown by SDS–PAGE analysis (Supplementary Fig. 1¹), were then pooled and concentrated using a Millipore Amicon Ultra 15 centrifugal filter (30 000 Da molecular-weight cutoff) followed by buffer exchange into 50 mM Tris–HCl, 2 mM DTT pH 7.5 using a PD-10 desalting column (GE Healthcare). The concentration of purified CmlS was determined by a Bradford assay (Bradford, 1976). Aliquots of CmlS were then flash-frozen in liquid nitrogen and stored at 193 K. We did not find it necessary to remove the N-terminal hexahistidine tag on recombinant CmlS to obtain high-quality crystals (see below).

The selenomethionine derivative of CmlS was produced in the methionine-auxotroph *E. coli* strain DL41(DE3) grown in M9

SeMET High Yield medium (Medicilon) using the same growth conditions as used for native protein expression. The selenomethionine-labelled CmlS was purified using the same procedure as described above.

2.2. Crystallization

All crystallization experiments were performed at room temperature. For initial screening, CmlS was kept at a concentration of 74 µM (~5 mg ml⁻¹) in buffer containing 50 mM Tris–HCl, 2 mM DTT pH 7.5. Sitting-drop vapour-diffusion trials were carried out using Qiagen crystallization screening kits in 96-well plates (Greiner): 1 µl protein solution was mixed with 1 µl crystallization solution and equilibrated

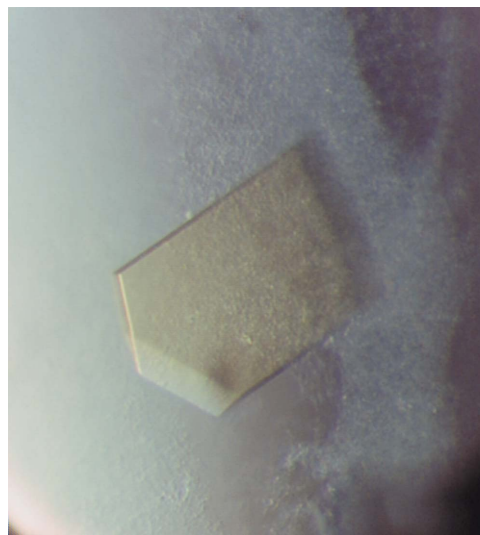


Figure 2
Optimized crystals of CmlS obtained using 0.1 M Na HEPES pH 6.8–7.4, 17–22% PEG 3350.

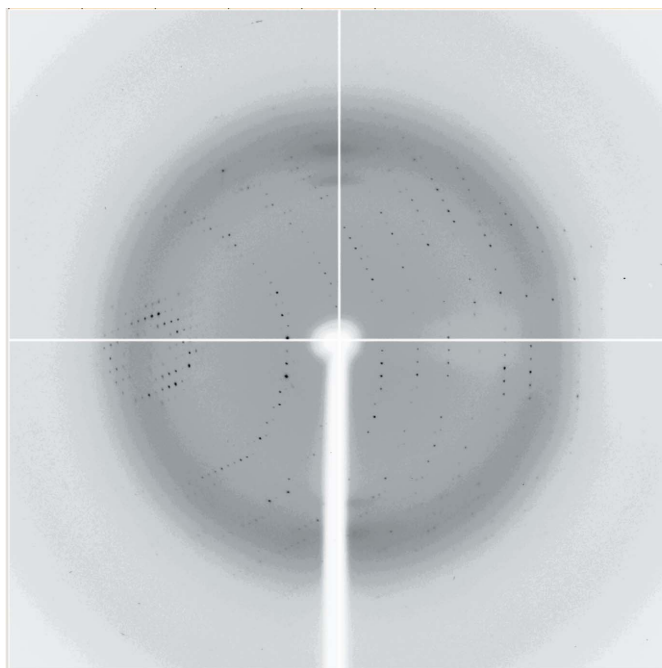


Figure 3
Diffraction of SeMet CmlS crystals. The data were collected to 2.0 Å resolution and processed to 2.2 Å resolution.

¹ Supplementary material has been deposited in the IUCr electronic archive (Reference: PU5246).

Table 1

Diffraction data for SeMet CmlS crystals.

Values in parentheses are for the highest resolution shell (2.28–2.20 Å).

Space group	C2
Unit-cell parameters (Å, °)	$a = 208.1, b = 57.7,$ $c = 59.9, \beta = 97.5$
Wavelength (Å)	0.98160
Temperature (K)	100
Resolution range (Å)	30–2.2 (2.28–2.20)
Observed reflections	762884
Unique reflections	36054
Data completeness (%)	99.7 (100)
Redundancy	7.2 (6.5)
R_{merge} (%) [†]	7.7 (5.5)
$\langle I/\sigma(I) \rangle$	35.7 (4.1)
Matthews coefficient (Å ³ Da ⁻¹)	2.68
Solvent content (%)	54.2

[†] $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ and $\langle I(hkl) \rangle$ represent the diffraction intensity values of the individual measurements and the corresponding mean values. The summation is over all measurements.

against 100 µl well solution. Initial hits were obtained in condition Nos. 54, 64 and 79 of the JCSG+ Suite (Qiagen), all of which contained PEG 3350 together with various salts and buffers.

Crystallization conditions were optimized using a grid screen of PEG 3350 *versus* the pH of sodium acetate buffer. 1 h prior to crystallization trials, a concentrated solution of L-arginine and L-glutamate at pH 7.6 was added to the CmlS sample to afford a final concentration of 50 mM of each amino acid. Optimization using the hanging-drop vapour-diffusion method was performed in 24-well plates (VDX): 2 µl CmlS solution was mixed with 2 µl crystallization solution on a siliconized glass cover slip and then equilibrated against 1 ml well solution. The final crystallization conditions contained 0.1 M Na HEPES pH 6.8–7.4, 17–22% PEG 3350. Crystals appeared after 3–4 d and reached their maximum size within a further 2 d. Typical crystal dimensions were 0.25 × 0.1 × 0.02 mm. Selenomethionine-derivative (SeMet) crystals of CmlS were obtained using the same procedure as used for the native CmlS crystals.

2.3. Data collection and diffraction measurements

Single-wavelength anomalous dispersion data were collected from SeMet CmlS crystals on the X12B beamline at Brookhaven National Laboratory National Synchrotron Light Source using an ADSC Quantum-4 CCD detector. All data were collected at 100 K. Prior to flash-freezing in liquid nitrogen, the CmlS crystals were sequentially immersed in crystallization solution containing 10, 15 and 20% PEG 200 as a cryoprotectant. Oscillations of 1° with an exposure time of 40 s per image and a crystal-to-detector distance of 200 mm were used. The data were indexed and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

3. Results and discussion

Recombinant CmlS was expressed in soluble form in *E. coli* using a low-temperature protocol. CmlS was then isolated in one step to achieve essentially single-band purity as confirmed by SDS-PAGE (Supplementary Fig. 1). The purified yield of CmlS was 32 mg per litre of culture. Concentrated CmlS in pure form had a distinct yellow hue characteristic of a flavin-containing enzyme. Accordingly, the UV-visible spectrum of CmlS revealed absorbance maxima (375 and 458 nm) typical of a bound oxidized flavin.

CmlS initially crystallized in condition Nos. 54, 64 and 79 of the JCSG+ Suite (Qiagen). Expansion of these conditions using the hanging-drop vapour-diffusion method in 24-well plates produced crystals that were large enough for diffraction studies (Fig. 2), but the initial diffraction was very weak. It has been reported that addition of 50 mM L-glutamate and 50 mM L-arginine greatly improves protein solubility and long-term stability (Golovanov *et al.*, 2004). The effect of these amino acids on crystal quality was assessed by adding them to CmlS samples immediately after thawing and at least 1 h prior to crystallization trials. The crystal morphology and size did not change beyond the normal range of variability; however, the diffraction limit of the crystals increased to 2.3 Å at the home source. The resulting large flawless crystals produced diffraction to 2.0 Å at the synchrotron (Fig. 3). A typical crystal had dimensions of 0.25 × 0.1 × 0.02 mm.

The crystals belonged to the C-centered monoclinic space group C2, with unit-cell parameters $a = 208.1, b = 57.7, c = 59.9$ Å, $\beta = 97.5^\circ$. Diffraction data for CmlS were processed in the resolution range 30–2.2 Å. A Matthews coefficient of 2.68 Å³ Da⁻¹ was obtained (Matthews, 1968), with a solvent content of 54.2%, representing the presence of a monomer in the asymmetric unit. A summary of the crystal parameters and the statistics of the diffraction data are presented in Table 1. Solution of the structure of CmlS is currently under way. In parallel, we are testing various compounds with CmlS (*e.g.* acetate, chloroacetate, malonate and thioesters of these compounds) in order to determine the optimal substrate for halogenation activity.

This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR). ZJ is a Canada Research Chair in Structural Biology. DLZ is the recipient of an Early Researcher Award.

References

- Blasiak, L. C. & Drennan, C. L. (2008). *Acc. Chem. Res.* doi:10.1021/ar800088r.
- Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
- Buedenbender, S., Rachid, S., Muller, R. & Schulz, G. E. (2008). *J. Mol. Biol.* doi:10.1016/j.jmb.2008.10.057.
- Dong, C., Flecks, S., Unversucht, S., Haupt, C., van Pee, K. H. & Naismith, J. H. (2005). *Science*, **309**, 2216–2219.
- Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A. & Gerwick, W. H. (2004). *Chem. Biol.* **11**, 817–833.
- Flecks, S., Patallo, E. P., Zhu, X., Ernyei, A. J., Seifert, G., Schneider, A., Dong, C., Naismith, J. H. & van Pee, K. H. (2008). *Angew. Chem. Int. Ed. Engl.* **47**, 9533–9536.
- Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. & Lian, L. Y. (2004). *J. Am. Chem. Soc.* **126**, 8933–8939.
- Gribble, G. W. (2003). *Chemosphere*, **52**, 289–297.
- Keller, S., Wage, T., Hohaus, K., Holzer, M., Eichhorn, E. & van Pee, K. H. (2000). *Angew. Chem. Int. Ed. Engl.* **39**, 2300–2302.
- Kirner, S., Hammer, P. E., Hill, D. S., Altmann, A., Fischer, I., Weislo, L. J., Lanahan, M., van Pee, K. H. & Ligon, J. M. (1998). *J. Bacteriol.* **180**, 1939–1943.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Neumann, C. S., Fujimori, D. G. & Walsh, C. T. (2008). *Chem. Biol.* **15**, 99–109.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Pirae, M., White, R. L. & Vining, L. C. (2004). *Microbiology*, **150**, 85–94.
- Winter, J. M., Moffitt, M. C., Zazopoulos, E., McAlpine, J. B., Dorrestein, P. C. & Moore, B. S. (2007). *J. Biol. Chem.* **282**, 16362–16368.
- Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. (2007). *Biochemistry*, **46**, 1284–1292.