Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The 2.7 Å resolution structure of the glycopeptide sulfotransferase Teg14

Matthew J. Bick,[a] Jacob J. Banik,[b] Seth A. Darst[a] and Sean F. Brady[b]*

[a]Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA, and [b]Howard Hughes Medical Institute, Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

Correspondence e-mail: sbrady@rockefeller.edu

Received 25 June 2010
Accepted 13 September 2010

PDB Reference: Teg14, 3nib.

The TEG gene cluster was recently isolated from an environmental DNA library and is predicted to encode the biosynthesis of a polysulfated glycopeptide congener. Three closely related sulfotransferases found in the TEG gene cluster (Teg12, Teg13 and Teg14) have been shown to sulfate the teicoplanin aglycone at three unique sites. Crystal structures of the first sulfotransferase from the TEG cluster, Teg12, in complex with the teicoplanin aglycone and its desulfated cosubstrate PAP have recently been reported [Bick *et al.* (2010), *Biochemistry*, **49**, 4159–4168]. Here, the 2.7 Å resolution crystal structure of the apo form of Teg14 is reported. Teg14 sulfates the hydroxyphenylglycine at position 4 in the teicoplanin aglycone. The Teg14 structure is discussed and is compared with those of other bacterial 3′-phosphoadenosine 5′-phosphosulfate-dependent sulfotransferases.

1. Introduction

Vancomycin and teicoplanin are natural product antibiotics used clinically to treat Gram-positive bacterial infections, including methicillin-resistant *Staphylococcus aureus* (MRSA; Nicolaou *et al.*, 1999; Wolter *et al.*, 2007). These traditional antibiotics of last resort are members of a large class of naturally occurring antibiotics known as glycopeptides. Almost 200 unique glycosylated, halogenated, acylated, alkylated and sulfated glycopeptide congeners have been characterized from examining the metabolites produced by cultured actinomycetes. In recent years, the discovery of structurally novel glycopeptide congeners from cultured bacteria has slowed significantly. The vast majority of bacteria present in nature remain recalcitrant to culturing and therefore represent a potentially rich source of additional glycopeptide congeners (Rappe & Giovannoni, 2003). Metagenomics, which relies on cloning DNA directly from environmental samples to access the genomic DNA of uncultured bacteria, provides a strategy for studying natural products generated by microbes that are not readily grown in pure culture. In a survey of DNA extracted from desert soil collected in Utah, we uncovered a biosynthetic gene cluster [the teicoplanin-like eDNA-derived (TEG) gene cluster] that is predicted to encode the biosynthesis of the first polysulfated glycopeptide congeners (Banik & Brady, 2008).

During their biosynthesis, unique glycopeptide congeners are functionalized by pathway-specific sets of glycosyltransferases, acyltransferases, alkyltransferases and sulfotransferases (Nicolaou *et al.*, 1999; Wolter *et al.*, 2007). The TEG gene cluster contains three closely related sulfotransferases (Teg12, Teg13 and Teg14) that have been shown to sulfate the teicoplanin aglycone at three unique sites

(Fig. 1). Glycosylated, acylated and methylated glycopeptides are quite common; however, no polysulfated congeners had been described prior to the cloning of the eDNA-derived TEG gene cluster. Although anionic glycopeptides have rarely been reported as natural products, the semisynthetic phosphono congener telavancin recently received FDA approval (Higgins *et al.*, 2005; Stryjewski *et al.*, 2005). The enzymatic synthesis of anionic glycopeptides using recombinant sulfotransferases could provide a facile means to access additional members of this rare class of congeners.

In a recent paper, we presented crystal structures of Teg12 in its apo form, Teg12 in a binary complex bound to the teicoplanin aglycone and Teg12 in a ternary complex bound to the teicoplanin aglycone and the desulfated cosubstrate 3′-phosphoadenosine-5′-phosphate (PAP). These studies provided insight into the sulfotransferase mechanism employed by this class of finishing enzymes. In the current paper, we present an apo structure of a second sulfotransferase from the TEG pathway, Teg14. While Teg12 sulfates the dihydroxyphenylglycine at position 3 in the teicoplanin aglycone, Teg14 sulfates the hydroxyphenylglycine at position 4. Teg14 displays many of the same structural characteristics found in other members of this class of enzymes. We discuss the Teg14 structure itself and compare and contrast this apo structure with those of other sulfotransferases.

2. Materials and methods

2.1. Cloning Teg14 from an environmental DNA library

Teg14 (Sulf3) was PCR-amplified (30 cycles of 368 K for 30 s, 333 K for 30 s and 345 K for 90 s; FailSafe system from Epicentre) from eDNA clone D30 with the following forward and reverse primers: Sulf3For (*Bcl*I), **GCGCTGATCA**ATGAACGGTATTCGATGGATC, and Sulf3Rev (*Hin*dIII), **GCGCAAGCTT**ACAATCCGCCCGTTAGCCGGC. The amplified product was doubly digested with *Bcl*I/*Hin*dIII and ligated into *Bam*HI/*Hin*dIII-digested pET28a (Banik & Brady, 2008).

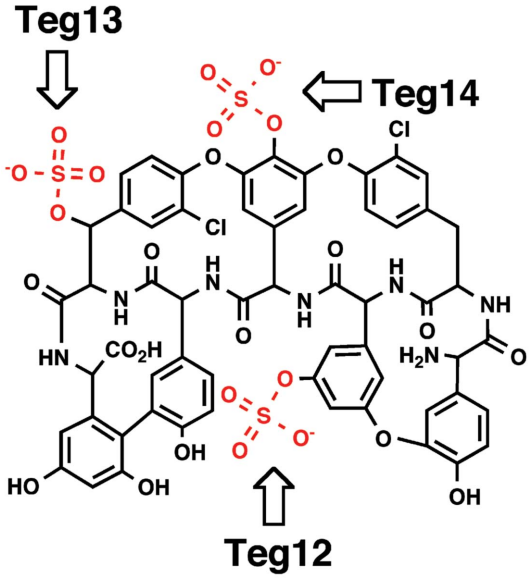

Figure 1
Trisulfated teicoplanin aglycone. The TEG sulfotransferases (Teg12, Teg13 and Teg14) sulfate three different sites on the teicoplanin aglycone.

2.2. Protein expression and purification

pET28a-Teg14 was transformed into *Escherichia coli* BL21 (DE3) for protein expression. Small-scale overnight cultures were used to inoculate (1:1000 dilution) 1 l cultures of LB plus 50 µg ml^{-1} kanamycin. The cultures were grown at 310 K until the density reached an OD_{600} of 0.6, at which point the temperature was reduced to 293 K. After 1 h, protein expression was induced with 0.5 m*M* IPTG. After 14–16 h, the cultures were harvested by centrifugation (3200*g* for 30 min). The cell pellet was resuspended in 40 ml lysis buffer [50 m*M* HEPES pH 7.5, 0.5 *M* NaCl, 5%(*v*/*v*) glycerol, 20 m*M* imidazole pH 8, 10 m*M* *β*-mercaptoethanol and 0.5% Triton X-100] and the cells were lysed by sonication. Crude cell lysates were centrifuged at 25 000*g* for 30 min and the supernatants were then incubated for 15 min (287 K) with 1 ml Ni–NTA resin (equilibrated with lysis buffer). After 15 min, the slurry was loaded onto a column and washed with 40 ml lysis buffer followed by 40 ml wash buffer [50 m*M* HEPES pH 7.5, 0.5 *M* NaCl, 5%(*v*/*v*) glycerol, 20 m*M* imidazole pH 8 and 10 m*M* *β*-mercaptoethanol]; the bound protein was eluted with 15 ml elution buffer [50 m*M* HEPES pH 7.5, 0.5 *M* NaCl, 5%(*v*/*v*) glycerol, 125 m*M* imidazole pH 8 and 10 m*M* *β*-mercaptoethanol]. The eluted protein was collected as seven ∼2 ml fractions. The 2–3 fractions with the highest protein concentration, as determined by the absorbance at 280 nm, were pooled for use in crystallization. This protein was concentrated to 20 mg ml^{-1} and buffer-exchanged three times into crystallization buffer [20 m*M* HEPES pH 7.5, 5%(*v*/*v*) glycerol, 200 m*M* NaCl, 5 m*M* DTT] using a 20 ml 30 000 molecular-weight cutoff Vivaspin concentrator (Vivascience). No attempt was made to remove the N-terminal His_6 tag, resulting in an additional 34 vector-derived amino acids appended to the N-terminus of the Teg14 protein sequence.

2.3. Crystallization and X-ray diffraction

Teg14 was screened at 20 mg ml^{-1} (295 K) for crystallization using the JCSG suites (Cores I–IV) from Qiagen. All screening was performed in 24-well trays (Qiagen) using the hanging-drop method. Crystallization drops consisted of 2 µl of a 1:1 mixture of protein and precipitant solutions and were equilibrated over a 500 µl reservoir volume. Three crystals appeared overnight in JCSG Core II condition A3 (1.0 *M* sodium citrate, 0.1 *M* CHES pH 9.5) and grew to maximum dimensions of approximately 300 × 300 × 300 µm in 4 d. Diffraction quality from two of the crystals was impaired while searching for suitable cryoconditions. The third crystal, which was harvested from the crystallization drop using a cryoloop (Hampton Research), briefly (<1 min) soaked in reservoir solution plus 30% glycerol and flash-cooled directly in a liquid-nitrogen cryostream at 100 K, diffracted to well beyond 2.5 Å resolution. These crystals proved to be irreproducible.

Diffraction data from the Teg14 crystal were measured using an R-AXIS IV^{++} area detector connected to a Rigaku/MSC MicroMax-007 HF Cu rotating-anode generator (wavelength 1.5418 Å) equipped with Varimax confocal Max-Flux optics at The Rockefeller University. Diffraction data were processed and scaled to 2.7 Å resolution using the *HKL*-2000 package (Otwinowski & Minor, 1997). The Teg14 crystal belonged to the trigonal space group $P3_121$, with unit-cell parameters $a = b = 112.19$, $c = 74.81$ Å, $\alpha = \beta = 90.0$, $\gamma = 120.0^\circ$. A single Teg14 monomer, including the 34 vector-derived residues, two molecules of glycerol and a single molecule of CHES (*N*-cyclohexyl-2-aminoethanesulfonic acid) buffer (total molecular weight of 34 822 Da) in the asymmetric unit gave a packing density V_M (Matthews, 1968) of 3.90 Å^3 Da^{-1}, corresponding to a solvent content of 68.43%. Diffraction data statistics are given in Table 1.

2.4. Structure determination and refinement

The Teg14 structure was solved by molecular replacement using the program *Phaser* (McCoy *et al.*, 2007; Storoni *et al.*, 2004). The Teg12 ternary complex structure (Bick *et al.*, 2010), devoid of all flexible loops and small-molecule substrates, was used as the search model. The ternary complex structure was chosen as the search model because it is the highest resolution structure from the group of three Teg12 structures reported. Teg12 has a sequence identity of 82.9% to Teg14. *Phaser* gave a single clear solution with a rotational *Z* score of 13.3, a translational *Z* score of 33.0 and a log-likelihood gain of 1278. No clashes were observed in the solution. A single Teg14 monomer was identified in the asymmetric unit. Rigid-body refinement of the molecular-replacement solution using *phenix.refine* (Adams *et al.*, 2010) gave an R_{work} and an R_{free} of 0.410 and 0.415, respectively. Manual rebuilding of the model using the resulting $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps was carried out in *Coot* (Emsley & Cowtan, 2004). Multiple rounds of restrained refinement were carried out using *phenix.refine*, employing the translation–libration–screw (TLS) protocol and two atomic displacement parameter (ADP) groups per residue (Grosse-Kunstleve & Adams, 2002; Winn *et al.*, 2001). Teg14 was partitioned into 15 TLS groups as follows: group 1, residues 0–12, 83–90, 57, 67 and 157–165 (the β-sheet core); group 2, residues 166–170; group 3, residues 171–186; group 4, residues 187–203; group 5, residues 204–214; group 6, residues 13–30 and the active-site CHES; group 7, residues 31–42; group 8, residues 43–56; group 9, residues 68–82; group 10, residues 91–101; group 11, residues 109–127; group 12, residues 133–156; group 13, residues 241–248; group 14, residues 249–261; group 15, residues 262–275 (Fig. 2). Side-chain atoms for which there was no observable density were removed from the model. After all protein residues had been assigned, the water-picking protocol in *phenix.refine* was used to identify density corresponding to waters. Several water molecules were then added manually. *Coot* identified five of these waters as 'unusual';

Table 1
Data-collection and refinement statistics for Teg14.

Values in parentheses are for the highest resolution shell.

Data collection	
Space group	$P3_121$
Unit-cell parameters (Å, °)	$a = b = 112.194$, $c = 74.807$, $\alpha = \beta = 90.0$, $\gamma = 120.0$
Temperature (K)	100
Matthews coefficient (Å^3 Da^{-1})	3.89
Solvent content (%)	68.43
Resolution (Å)	25.00–2.70 (2.80–2.70)
Wavelength (Å)	1.5418
R_{merge} (%)	0.073 (0.494)
$\langle I/\sigma(I)\rangle$	14.6 (2.5)
Completeness (%)	95.8 (93.1)
Reflections (observed)	39851
Reflections (unique)	14590
Redundancy	2.7 (2.2)
Refinement	
Resolution (Å)	22.79–2.70
Reflections	13643
R_{work}/R_{free} (%)	19.55/23.56
No. of monomers in asymmetric unit	1
No. of non-H atoms	
Total	1876
Protein	1797
Ligand	25
Water	54
B factors (Å^2)	
Overall	59.31
Protein	59.19
Ligand	73.02
Water	57.19
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.805
Dihedral angles (°)	14.59
MolProbity validation	
Clashscore (all-atom contacts)	20.8 [82nd percentile]
Poor rotamers (%)	0.57 [Arg127]
Ramachandran outliers (%)	0.00
Ramachandran favored (%)	97.39
C^β deviations > 0.25 Å	0
MolProbity score	1.93 [98th percentile]
Residues with bad bonds (%)	0.00
Residues with bad angles (%)	0.42 [Pro58]

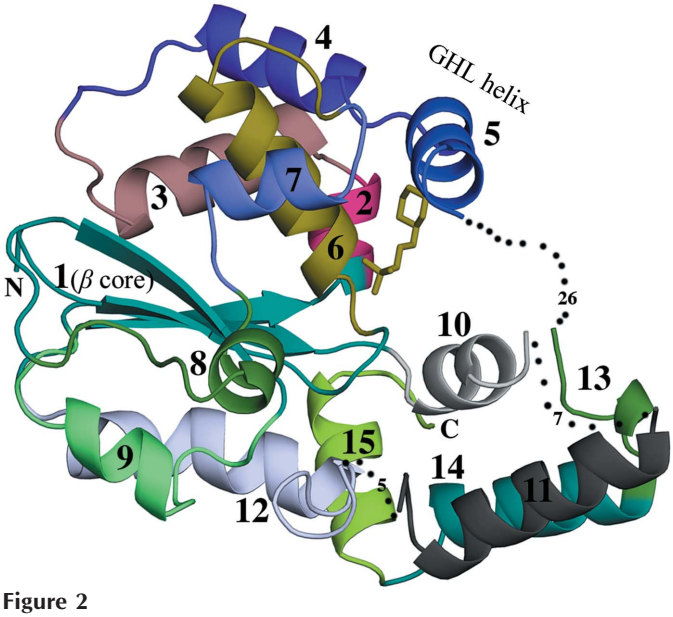

Figure 2
Teg14 colored according to the groups used for TLS refinement in *phenix.refine*. The groups are numbered from 1 to 15 (see §2 for the residues for each group). Regions of disorder are connected with gray dots and the number of disordered residues is indicated.

however, manual inspection suggested that they were reasonably assigned. Both X-ray to stereochemistry ('optimize wxc') and X-ray to ADP ('optimize wxu') weights were optimized during the final round of refinement. The model converged to a final R_{work} and R_{free} of 19.55% and 23.56%, respectively. The structure was validated using the *MolProbity* server at Duke University (Chen *et al.*, 2010), the results of which, together with refinement statistics, are presented in Table 1. All figures of the protein model were generated using the program *PyMOL* (DeLano, 2002).

3. Results

3.1. Overall structure and comparison with Teg12 and StaL

N-terminally His_6-tagged Teg14 was affinity-purified using nickel–NTA resin and crystallized without the need for further

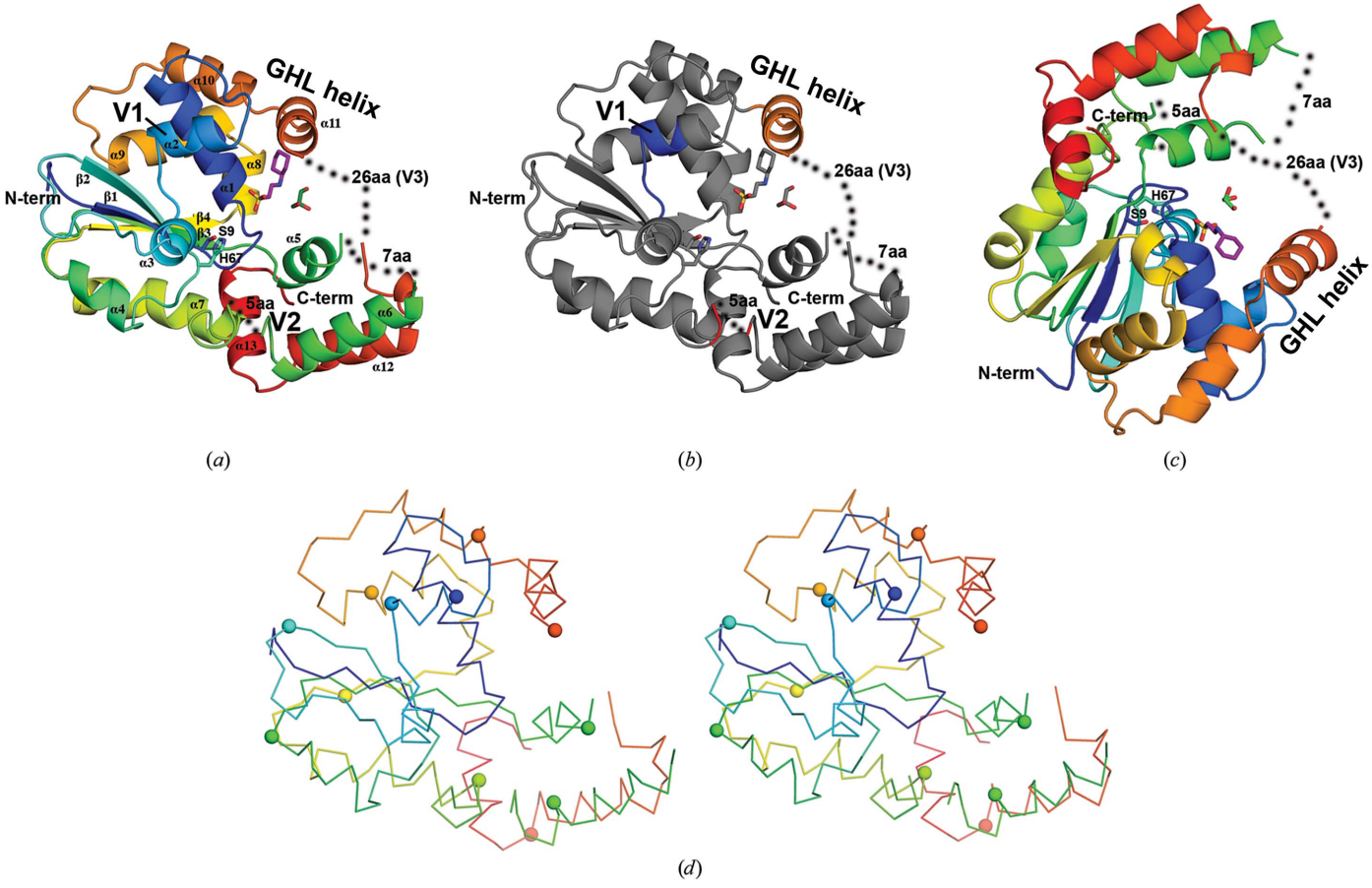

Figure 3
Overview of the 2.7 Å resolution crystal structure of Teg14. The model is colored from blue to red from the N- to the C-terminus. Regions of disorder are represented by gray dots, with the number of disordered residues indicated. (*b*) The same view as (*a*), but with the protein colored gray in order to accentuate the regions of sequence variability V1 (blue) and V2 (red). V3, which is encompassed by the 26-amino-acid disordered region, was not visible in the structure. Also shown in orange is the GHL helix, which adopts the bent conformation as seen in the Teg12 ternary complex and StaL structures. (*c*) An alternate view of (*a*). The molecules of CHES and glycerol in the active site are shown as sticks. (*d*) Stereoview of a C^{α} trace of Teg14. Every 20th C^{α} atom is shown as a sphere (residue 220 is not shown as it is disordered in our structure).

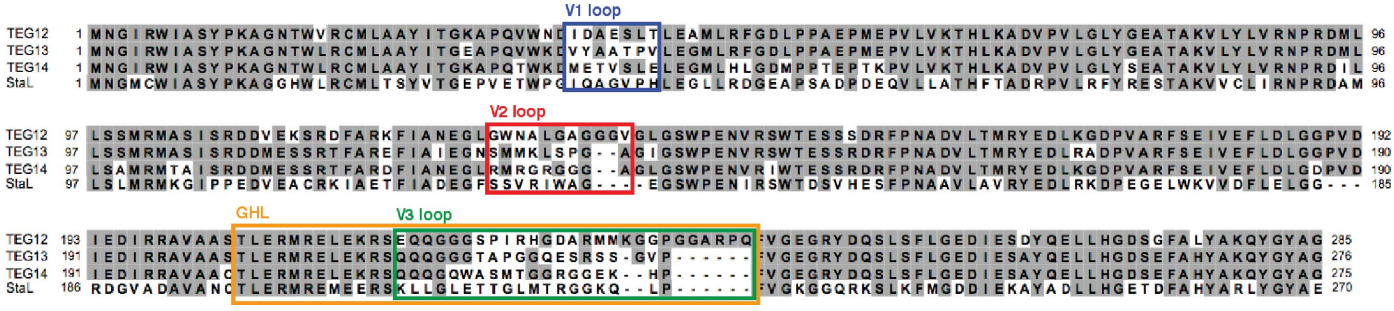

Figure 4
ClustalW sequence alignment of glycopeptide sulfotransferases. The three highly variable regions V1, V2 and V3 as well as the GHL are marked with different colored boxes.

purification. The structure was solved to 2.7 Å resolution by molecular replacement using the Teg12 ternary structure as a search model (PDB code 3nib; Fig. 3). The overall structure (Figs. 4 and 5) closely resembles those of both Teg12 (82.9% sequence identity) and StaL (52.2% sequence identity), a related sulfotransferase that is involved in the biosynthesis of the monosulfated glycopeptide congener A47934 (Shi *et al.*, 2007). Sulfotransferases from this family are biologically active as dimers. Teg14 dimerizes in the same manner as Teg12 and StaL. The Teg14 crystallographic asymmetric unit is comprised of a single monomer and the dimer is reconstituted by crystallographic symmetry. The dimer interface consists of a symmetrical interaction between a short helix–loop motif from one monomer and the same helix–loop motif from the other monomer. As seen in Teg12, there is a key hydrogen bond between the carbonyl O atom of Gly51 at the end of the helix from one monomer and the backbone N atom of Val74 of the other monomer. The side chain of Glu65 from one

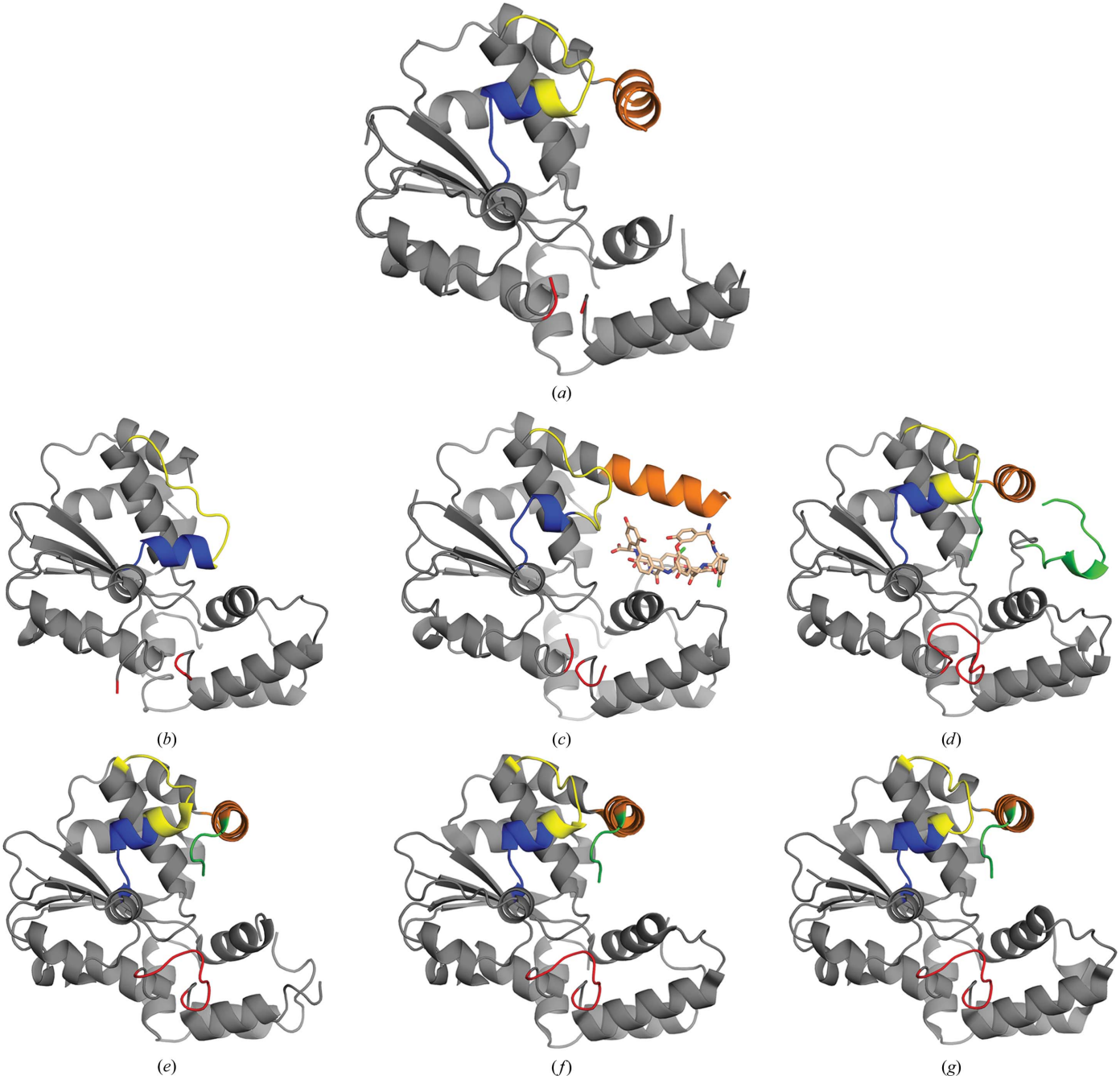

Figure 5
Comparison of the Teg12, Teg14 and StaL crystal structures. (*a*) Teg14, (*b*) apo Teg12, (*c*) Teg12 in a binary complex with the teicoplanin aglycone, (*d*) Teg12 in a ternary complex with PAP and the teicoplanin aglycone, (*e*) apo StaL, (*f*) StaL complexed with PAP and (*g*) StaL–sulfate. Regions of sequence variability are colored blue (V1), red (V2) and green (V3). The flexible GHL helix is colored orange. An additional region (residues 28–36) that shows structural variability among the Teg14, Teg12 and StaL structures is colored yellow. This region is contiguous with V1 (residues 37–43).

monomer is also within hydrogen-bonding distance of the side chains of Thr66 and Lys65 of the opposite monomer. Additional hydrophobic contacts that exist between the two interface helices help to stabilize the Teg14 dimer. In the Teg14 dimer, the active-site cavity of each monomer faces the dimer interface.

As with other sulfotransferases in this family, Teg14 consists of a single α/β globular domain composed of a parallel β-sheet core surrounded by α-helices (Fig. 3). The β-sheet core contains four strands. A *ClustalW* alignment of the TEG sulfotransferases shows that almost all of the variability that exists between these three enzymes resides in three highly variable sequences: V1, V2 and V3 (Fig. 4). In Teg14, these variable regions correspond to Met37–Glu43 (V1), Arg127–Gly134 (V2) and Gln215–Pro233 (V3). All three variable regions are conformationally flexible loops that we have hypothesized to play key roles in determining both the substrate specificities and regiospecificities of the TEG sulfotransferases (Bick *et al.*, 2010).

Gly28–Glu43, which encompass V1, adopts three highly divergent conformations in the Teg12 structures. The observed conformation of this section in the current Teg14 apo structure, in which a large proportion of this region is helical (Trp34–Val40), is likely to be preferred as this conformation is also seen in the Teg12 ternary complex (Fig. 5*d*) and the three different StaL structures (Figs. 5*e*, 5*f* and 5*g*). Met128–Gly132 of V2 and a long section (Gln215–Asp240) that fully encompasses V3 could not be modeled in the electron-density map; these regions are presumed to be disordered. The V3 loop is part of a larger motif that we have termed the GHL (glycopeptide helix loop) and is believed to interact extensively with the glycopeptide. N-terminal to V3 is the helical portion of the GHL. In our Teg12 structures this region is either part of a single long helix (Ile193–Glu216) that associates with the glycopeptide substrate or two separate short helices that result from the unwinding (bending) of the long helix. In the Teg14 structure the GHL helix adopts the bent conformation observed in the Teg12 ternary structure and the StaL structures (Fig. 5). We have proposed that the winding and unwinding of the GHL helix may open and close a pocket into which the glycopeptide binds and could also help to direct the glycopeptide to the active site.

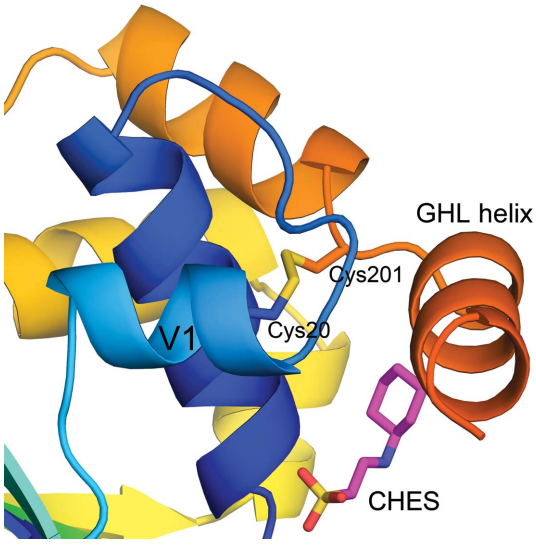

Figure 6
Teg14 shown in an orientation that highlights the disulfide bond formed by Cys20 and Cys201. The protein backbone is colored as a rainbow as in Fig. 3.

Cys201 of the GHL helix makes a disulfide cross-link to Cys20 of helix 1 (Fig. 6). A disulfide is also present at this position in StaL (Cys20–Cys196) but is not present in Teg12, in which Cys201 corresponds to Ser203. The Cys201–Cys20 disulfide occurs at the point where the GHL helix bends in most sulfotransferases. In the Teg12 binary structure (Fig. 5*c*) the GHL helix adopts a single straight conformation, while in other glycopeptide sulfotransferase structures (including this Teg14 structure) it either bends or becomes disordered at Cys201 (Fig. 5). The portion of the GHL helix that displays this conformational diversity is directly C-terminal to Cys201. The lack of a disulfide in Teg12 could explain why the region C-terminal to Ser203 is disordered in the Teg12 apo structure but not in the Teg14 or StaL structures. At this point it is still unclear whether the presence or absence of the disulfide plays a functional role in helping to control the bending of the GHL.

3.2. The Teg14 active site

Teg14 was crystallized in its apo form. During refinement, two regions of prominent electron density were observed in the active-site cavity. One of these regions was easily assigned as a molecule of CHES, a component of the crystallization solution. The CHES molecule is oriented in much the same way as PAP is in the Teg12 ternary complex and StaL–PAP structures. The majority of the residues important for binding PAP also coordinate CHES (Fig. 7). Interactions between CHES and the protein are primarily mediated through van der Waals contacts, although there are several key hydrogen bonds. The hydroxyl O atom of Tyr165 hydrogen bonds to the N atom of CHES. This tyrosine is in the same general vicinity in the Teg12 and StaL structures (Tyr167 in Teg12 and Tyr163 in StaL), but is located too far from PAP to hydrogen bond to either the 3′-phosphate or the O atom of the ribose. Mutation of Tyr167 in Teg12 renders the protein insoluble, implicating that it plays a role in protein stability. Thr16 is also within hydrogen-bonding distance of the CHES sulfate (Figs. 6 and 7). The backbone N atoms of Ala13, Gly14, Asn15 and Thr16 of the 5′-phosphosulfate-binding motif (5′-PSB) and of Trp17 are all within contact distance of the sulfate from CHES.

The second region of electron density that was enhanced during refinement proved more difficult to assign. This density is in the immediate vicinity of the hydroxyl of Ser98, a residue which is part of the 3′-phosphate-binding motif (3′-PB) and is important for binding 3′-phosphoadenosine-5′-phosphosulfate (PAPS). Several small molecules and ions were modeled into the density and subjected to refinement in order to identify the best fit. We initially thought that sulfate could be correct; however, when sulfate was placed in the model and refined the resulting *B* factors for the sulfate atoms were in the range of

140 Å^2. The *B* factor for the adjacent side-chain O atom of Ser98 is 90.5 Å^2. Imidazole was also tried and gave *B* factors around 90 Å^2, but did not fit the density well. A chloride ion gave the lowest *B* factor (82 Å^2), but did not satisfy all of the density. Ultimately, glycerol was chosen as the ligand that best fitted the density and gave refined *B* factors of around 90 Å^2 (Fig. 7*a*). The StaL–sulfate structure has a sulfate modeled in the location where we have placed glycerol. Only three residues make significant contacts with the putative glycerol: Lys12, Arg90 and Ser98. Of these, only Ser98 is within hydrogen-bonding distance. A large portion of the glycerol is exposed to the solvent region. This fact could explain the high temperature factors observed for glycerol. Although a water molecule is within 3.3 Å of a glycerol O atom, the solvent overall in this region is poorly ordered.

It is clear from our structural studies of Teg12 and Teg14 that the PAPS cosubstrate-binding pocket can bind a number of molecules other than PAPS and PAP. The apo Teg12 structure contains a molecule of the dipeptide aspartame in the cosubstrate pocket (Fig. 7*c*). In the Teg14 structure the three separate compounds that fill the PAPS-binding pocket (CHES, glycerol and H_2O) appear to mimic the general PAP-binding motif seen in both the Teg12 and StaL binary complex structures. The cyclohexyl ring from CHES is bound in the same position and orientation as the adenine of PAP. It aligns parallel with Trp17, mimicking how the adenine stacks with Trp17 in Teg12 and StaL, and its upper edge is bound by Leu203 and Met206 of the GHL helix, the same two residues that interact with the adenine in the Teg12 ternary complex (Leu205 and Met208) and StaL–PAP (Leu198 and Met201) structures. The sulfonic acid from CHES fills the 5′-phosphate-binding pocket, the glycerol fills the 3′-phosphate pocket and the water is positioned in the site predicted to bind the sulfate of PAPS. Even though CHES, aspartame and PAPS are structurally distinct from one another, all three bind the PAPS-binding pocket in a similar fashion (Fig. 7).

Teg-like sulfotransferases are thought to catalyze the sulfonation reaction through an in-line attack mechanism by utilizing histidine as a general base (Chapman *et al.*, 2004; Negishi *et al.*, 2001). In the Teg14 structure, His67 is positioned approximately 7 Å from the sulfate of the ordered molecule of CHES that is in the PAPS-binding pocket (Fig. 8). A well ordered water molecule occupies the intermediate space where the sulfate from PAPS would be and is within hydrogen-bonding distance of His67. Water also occupies this space in the Teg12 binary and ternary complex structures and in the apo StaL structure. A sulfate takes the place of the His67 water in the StaL–sulfate structure. In all three Teg12 struc-

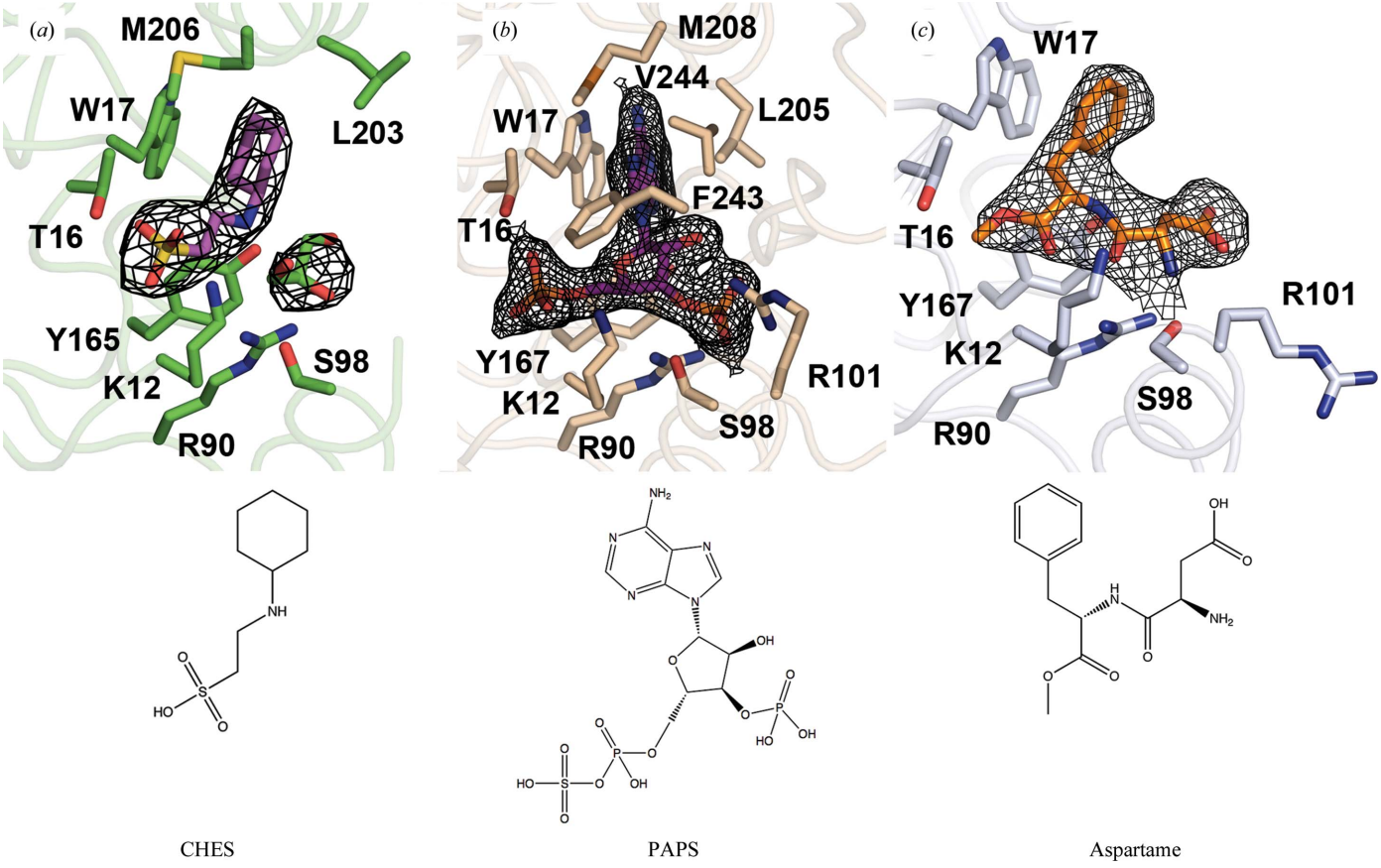

Figure 7
A comparison of the Teg14 (*a*), Teg12 ternary complex (*b*) and apo Teg12 (*c*) active sites. Several of the key residues that contact PAP in the Teg12 structure and aspartame in the apo Teg12 structure also coordinate CHES in Teg14. Sections of the $2|F_o| - |F_c|$ map surrounding the active-site ligands (CHES and glycerol in Teg14, PAP in Teg12 and aspartame in apo Teg12) are shown contoured at 1σ.

tures, all three StaL structures and the Teg14 structure, the basic N^{ε} atom of the imidazole ring from His67 is positioned facing Ser9 and within hydrogen-bonding distance.

4. Discussion

The TEG gene cluster is predicted to encode the biosynthesis of a heptapeptide (Hpg-Bht-Dpg-Hpg-Hpg-Bht-Dpg, where Hpg, Bht and Dpg are hydroxyphenylglycine, β-hydroxytyrosine and dihydroxyphenylglycine, respectively) that is oxidatively cross-linked into the same four macrocycles. This oxidatively cross-linked heptapeptide skeleton only differs from the teicoplanin aglycone by the substitution of Bht for Tyr at the second position in the heptapeptide. We have shown that each TEG sulfotransferase can utilize the teicoplanin aglycone as a substrate (Banik & Brady, 2008). Teg12, Teg13 and Teg14 sulfate the teicoplanin aglycone at different locations to produce three unique monosulfated glycopeptide derivatives. In combination, two Teg enzymes can be used to produce three different disulfated teicoplanins and when all three Teg enzymes are used in conjunction a trisulfated teicoplanin is produced (Fig. 1). We initially embarked on structural studies of the TEG-pathway sulfotransferases not only to improve our understanding of the sulfation mechanism, but also with the idea of exploring the possibility of engineering this group of enzymes to accept a broader range of glycopeptide scaffolds.

Attempts to obtain cocrystals of Teg14 with the teicoplanin aglycone, the monosulfated teicoplanin products of Teg12 and Teg13 or the disulfated teicoplanin product of Teg12 and Teg13 have been unsuccessful. Therefore, at present we do not know the precise orientation of the glycopeptide substrate in the Teg14 active site. In the Teg12 binary complex structure, the hydroxyl to be sulfated from residue 3 of the glycopeptide is bound more than 16 Å from the catalytic histidine and the modeled PAPS-associated sulfate. At some point in the reaction cycle, the glycopeptide must reorient from its observed position in the Teg12 binary complex structure to a position where the hydroxyl to be sulfated is in direct proximity of the active-site His67 and PAPS-associated sulfate. Such a conformational rearrangement would require the glycopeptide-binding pocket to exhibit a high degree of conformational flexibility. The existence of a conformationally flexible substrate-binding pocket in this family of sulfotransferases is supported by the fact that the three regions predicted to interact with the glycopeptide (V1–3) are largely disordered loops in all three Teg12 structures as well as in the Teg14 structure reported here. It is likely that these large conformationally flexible substrate-binding pockets would accept or through mutagenesis could be engineered to accept a range of glycopeptide scaffolds other than the teicoplanin aglycone used in our studies.

The Teg14 structure provides another example of how the GHL helix behaves in the absence of a substrate. The GHL helix adopts the bent conformation seen in Teg14 in all available structures of glycopeptide sulfotransferases in which no substrate is present in the active site. The most dramatic conformational difference seen within available glycopeptide sulfotranferase structures is observed in the Teg12 binary complex structure which is complexed with the teicoplanin aglycone. In this structure the GHL helix adopts a markedly different conformation from that seen in all other structures (Fig. 5*c*). Outside of the GHL helix (Fig. 5, orange), the regions of StaL, Teg12 and Teg14 that are ordered look essentially the same in all structures. Structural differences in the V1 region are subtle and, especially for the apo Teg12 structure, may be crystallographically induced (Fig. 5*b*, yellow).

Each TEG sulfotransferase adds a sulfate to a different location on the glycopeptide substrate. The glycopeptide must therefore be arranged differently in each active site in order for the in-line sulfuryl-transfer reaction to proceed on three separate residues. Although we have proposed that the GHL helix may help to bring teicoplanin into close proximity of His67 through unwinding of the helix from the straight to the bent conformation, this helix is probably not responsible for positioning the glycopeptide in three different orientations in the active site, as its sequence is invariant amongst the Tegs and nearly so in StaL (Fig. 4). Instead, the V1, V2 and V3 loops, which differ significantly in each TEG sulfotransferase, are likely to play key roles in both recruitment and orientation of the glycopeptide into the active site. Additional structures of TEG sulfotransferases complexed with glycopeptides positioned within the active-site cavity are needed to confirm this hypothesis.

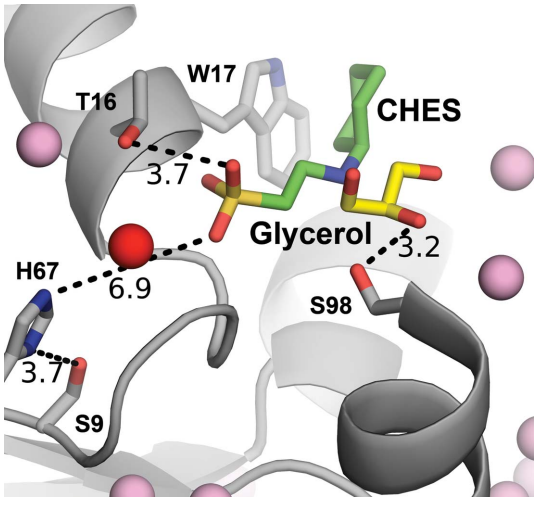

Figure 8
The orientation of CHES relative to the proposed Teg14 active-site residues His67 and Ser9. The side chain of His67 is approximately 7 Å from the sulfate of CHES. Water molecules are shown as pink spheres. The water molecule in the position where the sulfate of PAPS would be is colored red.

We thank Deena Oren of The Rockefeller University Structural Biology Resource Center for help with data collection. This work was supported by National Institutes of Health grant GM077516, the Howard Hughes Medical Insti-

tute and the Beckman Foundation. The use of the Rigaku/MSC microMax 007HF in the Rockefeller University Structural Biology Resource Center was made possible by Grant Number 1S10RR022321-01 from the National Center for Research Resources of the NIH.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Banik, J. J. & Brady, S. F. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 17273–17277.
Bick, M. J., Banik, J. J., Darst, S. A. & Brady, S. F. (2010). *Biochemistry*, **49**, 4159–4168.
Chapman, E., Best, M. D., Hanson, S. R. & Wong, C. H. (2004). *Angew. Chem. Int. Ed. Engl.* **43**, 3526–3548.
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.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, California, USA.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Grosse-Kunstleve, R. W. & Adams, P. D. (2002). *J. Appl. Cryst.* **35**, 477–480.
Higgins, D. L., Chang, R., Debabov, D. V., Leung, J., Wu, T., Krause, K. M., Sandvik, E., Hubbard, J. M., Kaniga, K., Schmidt, D. E. Jr, Gao, Q., Cass, R. T., Karr, D. E., Benton, B. M. & Humphrey, P. P. (2005). *Antimicrob. Agents Chemother.* **49**, 1127–1134.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
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.
Negishi, M., Pedersen, L. G., Petrotchenko, E., Shevtsov, S., Gorokhov, A., Kakuta, Y. & Pedersen, L. C. (2001). *Arch. Biochem. Biophys.* **390**, 149–157.
Nicolaou, K. C., Boddy, C. N. C., Bräse, S. & Winssinger, N. (1999). *Angew. Chem. Int. Ed. Engl.* **38**, 2096–2152.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rappe, M. S. & Giovannoni, S. J. (2003). *Annu. Rev. Microbiol.* **57**, 369–394.
Shi, R., Lamb, S. S., Bhat, S., Sulea, T., Wright, G. D., Matte, A. & Cygler, M. (2007). *J. Biol. Chem.* **282**, 13073–13086.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Stryjewski, M. E., O'Riordan, W. D., Lau, W. K., Pien, F. D., Dunbar, L. M., Vallee, M., Fowler, V. G. Jr, Chu, V. H., Spencer, E., Barriere, S. L., Kitt, M. M., Cabell, C. H. & Corey, G. R. (2005). *Clin. Infect. Dis.* **40**, 1601–1607.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Wolter, F., Schoof, S. & Süssmuth, R. (2007). *Top. Curr. Chem.* **267**, 143–185.