Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Preliminary X-ray characterization of a novel type of anchoring cohesin from the cellulosome of *Ruminococcus flavefaciens*

Orly Alber,[a] Ilit Noach,[a] Raphael Lamed,[b,c] Linda J. W. Shimon,[d] Edward A. Bayer[a]* and Felix Frolow[b,c]*

[a]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel, [b]Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv 69978, Israel, [c]The Daniella Rich Institute for Structural Biology, Tel Aviv University, Tel Aviv 69978, Israel, and [d]Department of Chemical Research Support, The Weizmann Institute of Science, Rehovot 76100, Israel

Correspondence e-mail: ed.bayer@weizmann.ac.il, mbfrolow@post.tau.ac.il

Received 15 November 2007
Accepted 17 December 2007

Ruminococcus flavefaciens is an anaerobic bacterium that resides in the gastrointestinal tract of ruminants. It produces a highly organized multi-enzyme cellulosome complex that plays a key role in the degradation of plant cell walls. ScaE is one of the critical structural components of its cellulosome that serves to anchor the complex to the cell wall. The seleno-L-methionine-labelled derivative of the ScaE cohesin module has been cloned, expressed, purified and crystallized. The crystals belong to space group $C2$, with unit-cell parameters a = 155.6, b = 69.3, c = 93.0 Å, β = 123.4°, and contain four molecules in the asymmetric unit. Diffraction data were phased to 1.95 Å using the anomalous signal from the Se atoms.

1. Introduction

The cellulosome is an extracellular multisubunit complex of bacterial origin comprised of catalytic and noncatalytic subunits and is involved in the efficient degradation of crystalline cellulosic substrates and associated plant cell-wall polysaccharides (Shoham *et al.*, 1999). Cellulosomal subunits are composed of several functional modules. The main component is a noncatalytic polypeptide, termed scaffoldin (Bayer *et al.*, 1994), which is composed of multiple copies of cohesin modules, each of which interacts with its complementary dockerin module borne by the cellulosomal enzymes. The high-affinity cohesin–dockerin interaction is involved in the specific integration of polysaccharide hydrolases into the cellulosome complex and contributes to its structural stability (Bayer *et al.*, 2004). To date, cohesins have been phylogenetically distributed into three groups according to sequence homology: the type I cohesins (Beguin *et al.*, 1992), the type II cohesins (Leibovitz & Beguin, 1996) and the recently discovered type III cohesins (Ding *et al.*, 2001). The dockerins that interact with each cohesin type are, by definition, of the same type.

Determination of the three-dimensional structure by X-ray crystallography provides key insight into the type-specific cohesin–dockerin interaction and the functional assembly of the cellulosome. Several crystal structures of type I and type II cohesins have been determined to date. The structures of three different type I cohesins, two of them from the CipA scaffoldin of *Clostridium thermocellum* (Shimon *et al.*, 1997; Tavares *et al.*, 1997) and another from the CipC scaffoldin of *C. cellulolyticum* (Spinelli *et al.*, 2000), have been determined, all of which share a jelly-roll topology forming a flattened nine-stranded β-sandwich. The structures of three different type II cohesins have also been solved: the first cohesin from the ScaB scaffoldin of *Acetivibrio cellulolyticus* (Noach *et al.*, 2003), the 11th cohesin of the ScaA scaffoldin of *Bacteroides cellulosolvens* (Noach *et al.*, 2005) and the cohesin module from the cell-surface anchoring protein SdbA from *C. thermocellum* (Carvalho *et al.*, 2005). The type II cohesins have the same jelly-roll topology as the type I cohesins from *C. thermocellum* and *C. cellulolyticum* but contain several additional structural elements, *i.e.* an α-helix and two 'β-flaps' that disrupt the normal course of β-strands 4 and 8. The structures of type I and type II cohesin–dockerin complexes from *C. thermocellum* have also been solved (Carvalho *et al.*, 2003, 2007; Adams *et al.*, 2006).

Ruminococcus flavefaciens is a Gram-positive anaerobic cellulosome-producing cellulolytic bacterium that is usually found in the rumen and hindgut of ruminants and other herbivorous animals (Julliand *et al.*, 1999; Krause *et al.*, 1999; Nelson *et al.*, 2003; Flint, 1997; Wedekind *et al.*, 1988). The cellulosomal system of *R. flavefaciens* strain 17 was recently discovered to bear one of the most intricate complexes in the cellulosomal world. According to sequence analysis of the *sca* gene cluster, four structural cellulosome components were found; ScaA, ScaB, ScaC and ScaE. It was shown that the primary ScaA scaffoldin, which contains three cohesins, binds specifically to either Cel44A-type dockerins or to the C-terminal dockerin of a singular 'adaptor' scaffoldin ScaC (Fig. 1*a*; Rincon *et al.*, 2004). The ScaA dockerin module interacts with each of the seven cohesins comprising ScaB (Fig. 1*b*; Ding *et al.*, 2001; Rincon *et al.*, 2003). The C-terminus of ScaB includes an X module of unknown function that precedes a cryptic dockerin module. This X-dockerin dyad (XDoc) was shown to bind specifically to the N-terminal ScaE cohesin (Fig. 1*c*; Rincon *et al.*, 2005; Jindou *et al.*, 2006). Moreover, the C-terminus of ScaE contains a typical Gram-positive signal motif that mediates its covalent sortase-mediated attachment to the bacterial cell wall. ScaE thus plays a major role in anchoring the cellulosomal complex *via* its interaction with ScaB (Rincon *et al.*, 2005). The ScaE cohesin was also found to bind to a cotton-binding protein A (CttA), which contains two newly discovered carbohydrate-binding modules (CBMs). Like ScaB, CttA bears a similar XDoc dyad at its C-terminus (Fig. 1*d*; Rincon *et al.*, 2007). The cohesins of ScaA, ScaB and ScaE were found to be phylogenetically distinct from the previously described type I and type II cohesins (Ding *et al.*, 2001; Rincon *et al.*, 2005). They were therefore designated as a new group of cohesins: the type III cohesins. Here, we report the purification, crystallization and preliminary X-ray characterization of the type III ScaE cohesin from the *R. flavefaciens* 17 anchoring scaffoldin.

2. Materials and methods

2.1. Cloning, expression and purification

The DNA encoding the cohesin module from the *scaE* scaffoldin gene of *R. flavefaciens* (gi:148726262) was cloned into the pET28a expression vector (Novagen, Madison, Wisconsin, USA) together with a sequence encoding for a hexa-His tag attached to the 5′ end, using *Nco*I and *Xho*I restriction enzymes. The resultant plasmid was sequenced by capillary electrophoresis using a DNA analyzer (Applied Biosystems, Foster City, California, USA) and was transferred to *Escherichia coli* strain BL21(DE3)pLysS.

Expression of the seleno-L-methionine-labelled cohesin module (residues 30–211, SWISS-PROT accession No. Q4A3Y2) from *R. flavefaciens* ScaE scaffoldin (SeMet *Rf*-ScaECoh) was conducted according to the method described previously (Van Duyne *et al.*, 1993) with minor modifications. Transformed cells from a culture grown overnight in 1 ml Luria–Bertani broth containing 50 µg ml^{-1} kanamycin were isolated and resuspended in 1 ml M9 minimal medium supplemented with glucose (4 mg ml^{-1}) and 50 µg ml^{-1} kanamycin and added to 1 l of the same medium pre-incubated at 310 K. Incubation of the culture was continued at 310 K with shaking until the growth culture reached an OD_{600} of 0.6. At this point, SeMet was added to a final concentration of 50 µg ml^{-1} along with the following amino acids, which were added as solids: lysine hydrochloride (100 mg), threonine (100 mg), phenylalanine (100 mg), leucine (50 mg), isoleucine (50 mg) and valine (50 mg). After an additional 15 min of shaking, 0.1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and the culture was grown for an additional 13 h. The cells were harvested by centrifugation (1600*g* for 15 min) at 277 K and resuspended in 20 ml binding buffer [25 m*M* 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) adjusted to pH 7.0, 25 m*M* NaCl and 5 m*M* imidazole]. The suspension was kept on ice during sonication, after which cell debris was removed by centrifugation (20 000*g* at 277 K for 20 min). The expressed His-tagged protein was purified by metal-chelate affinity chromatography using an FPLC AKTA-prime (Amersham Pharmacia Biotech). The collected supernatant fractions were applied onto a column packed with 5 ml nickel–nitrilotriacetic acid resin (Amersham Pharmacia Biotech) equilibrated with binding buffer. Bound proteins were eluted in 1 ml fractions with binding buffer containing 150 m*M* imidazole; these fractions were further examined for protein purity by SDS–PAGE. The pooled purified fractions were then applied onto a Hi-Load 16/60 Superdex 75 size-exclusion column (Amersham Pharmacia Biosciences) equilibrated with 25 m*M* HEPES adjusted to pH 7.0, 25 m*M* NaCl and 1 m*M* dithiothreitol. The collected fractions (1 ml) were then pooled and concentrated using Vivaspin columns of 5000 Da molecular-weight cutoff (Vivascience), yielding 1.5 ml purified concentrated cohesin (11 mg ml^{-1}). The protein concentration was estimated by measuring the absorbance at 280 nm using the calculated extinction coefficient of the protein (ε_{280} = 18 450).

2.2. Mass-spectrometric analysis

Mass measurements were performed on an API 3000 Electrospray-Quadrupole tandem mass spectrometer (MDS-Sciex/ABI, Canada) equipped with a nanoelectrospray source (MDS Proteomics, Odense, Denmark). The molecular-mass difference between the SeMet and the native *Rf*-ScaECoh was determined to be 145 mass units. This value corresponds to an average content of 3.1 Se atoms per molecule of the labelled protein and is consistent with the presence of three Met residues in *Rf*-ScaECoh.

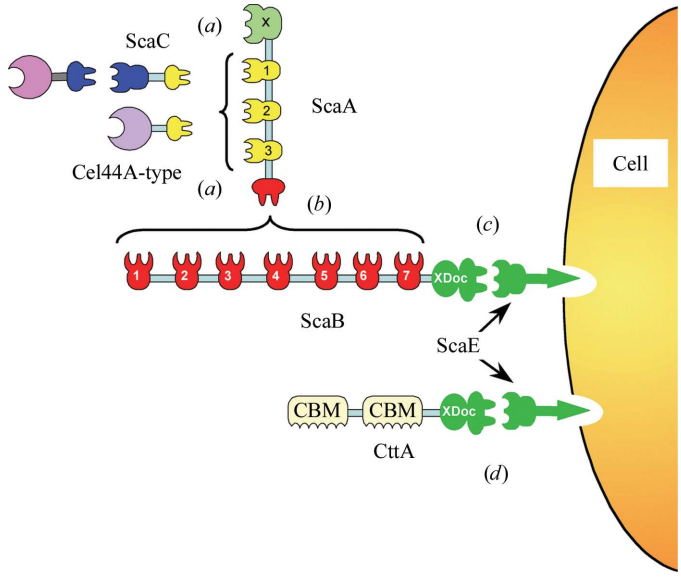

Figure 1
Schematic overview of the proposed modular interactions in the cellulosome system of *R. flavefaciens* strain 17. (*a*) The primary three-cohesin scaffoldin ScaA binds specifically to either Cel44A-type dockerins or to the C-terminal dockerin of ScaC. (*b*) The ScaA dockerin module interacts with one of the seven type III cohesins comprising the adaptor scaffoldin ScaB. (*c*) The conserved XDoc dyad of ScaB interacts with the cohesin module of the anchoring scaffoldin ScaE. (*d*) The ScaE cohesin also interacts with the XDoc dyad of the cellulose-binding protein CttA.

2.3. Crystallization

SeMet *Rf*-ScaECoh was crystallized at 293 K by the hanging-drop vapour-diffusion method. Screening for crystallization conditions was performed using the Hampton Research Index kit. Initial screening yielded single crystals using condition No. 42 (0.1 *M* bis-tris pH 5.5 and 25% PEG 3350). Further optimization of conditions yielded a final reservoir solution consisting of 0.1 *M* bis-Tris pH 5.5 and 23% PEG 3350. The protein solution (4 µl of a 11 mg ml^{-1} protein solution in 25 m*M* HEPES adjusted to pH 7.0, 25 m*M* NaCl and 1 m*M* dithiothreitol) was mixed with 6.8 µl reservoir solution and 1.2 µl additive solution (0.1 *M* calcium chloride; Additive Screen, Hampton Research) and equilibrated against 0.5 ml reservoir solution in a 24-well VDX plate (Hampton Research). The crystals grew within 2 d at 293 K and reached their full dimensions of 0.2 × 0.05 × 1 mm within 5 d, exhibiting plate morphology (Fig. 2).

2.4. X-ray data collection and processing

Diffraction data were collected from the SeMet *Rf*-ScaECoh crystals using synchrotron radiation (ESRF, Grenoble, France). Crystals for diffraction experiments were grown in the crystallization laboratory of the ID14 beamstation. The crystals were gently removed from the drop, incubated for a moment in a solution mimicking the mother liquor with the addition of 25% ethylene glycol and flash-frozen to 100 K at beamline ID23-1, which was equipped with an ADSC Q315 CCD area detector. Diffraction data were collected at the Se edge of 0.9792 Å, which was determined by a fluorescence energy scan of the mounted crystal. The crystals diffracted to a resolution of 1.95 Å. The diffraction images were taken at 0.5° oscillation steps over a 360° oscillation range (720 frames). Images were integrated and scaled using *DENZO* and *SCALEPACK* as implemented in *HKL*-2000 (Otwinowski & Minor, 1997) and additionally using *XDS* (Kabsch, 1993). The crystals belonged to the monoclinic space group *C*2, with unit-cell parameters a = 155.6, b = 69.3, c = 93.0 Å, β = 123.4°. Data-collection and processing statistics are summarized in Table 1.

The calculated Matthews coefficients of 3.35, 2.51 and 2.01 Å^3 Da^{-1} (Matthews, 1968) give solvent contents of 63, 51 and 38% and correspond to the presence of three, four and five monomers in the asymmetric unit, respectively. Pseudo-translation was not detected in the Patterson function map. The self-rotation function was calculated by *POLARRFN* as implemented in the *CCP*4 package (Collaborative Computational Project, Number 4, 1994) using data in the resolution range 41.5–2 Å. Four strong peaks representing noncrystallographic twofold axes were found on the κ = 180° section of the self-rotation function in polar coordinates (Fig. 3).

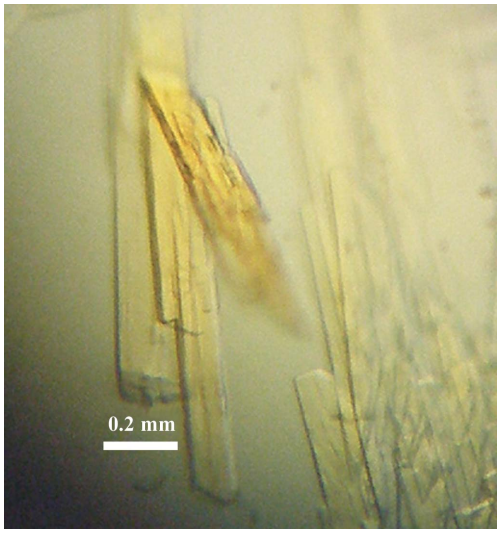

Figure 2
Crystals of SeMet *Rf*-ScaECoh (the N-terminal type III cohesin module from the anchoring scaffoldin ScaE) from *R. flavefaciens* strain 17.

Table 1
Data-collection and processing statistics for SeMet *Rf*-ScaECoh crystals.

Values in parentheses are for the highest resolution shell.

Experimental conditions	
X-ray source	ESRF, ID23-EH1
Wavelength (Å)	0.9792 (Se X-ray absorption peak)
Temperature (K)	100
Detector	ADSC Q315 CCD
Crystal parameters	
Space group	*C*2
Unit-cell parameters (Å, °)	a = 155.6, b = 69.3, c = 93.0, β = 123.4
Resolution (Å)	77.61–1.95 (1.98–1.95)
Mosaicity (°)	0.448
Solvent content (%)	51.01
Monomers in ASU	4
Data processing	
No. of measured reflections	407234
No. of unique reflections	60499
Completeness (%)	99.0 (89.9)
Mean $I/\sigma(I)$	29.9 (2.26)
$R_{\rm merge}$†	0.078 (0.372)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \overline{I(hkl)}|/\sum_{hkl}\sum_i I_i(hkl)$, where $\sum_{hkl}$ denotes the sum over all reflections and $\sum_i$ is the sum over all equivalent and symmetry-related reflections (Stout & Jensen, 1968).

2.5. Phasing

The structure was solved by SAD (single-wavelength anomalous diffraction) using the *SHELXC*/*D*/*E* sequence of programs (Sheldrick, 2008) as implemented in *CCP*4. Following the location of 17 heavy-atom sites (presumably Se) by *SHELXD*, primary phasing and

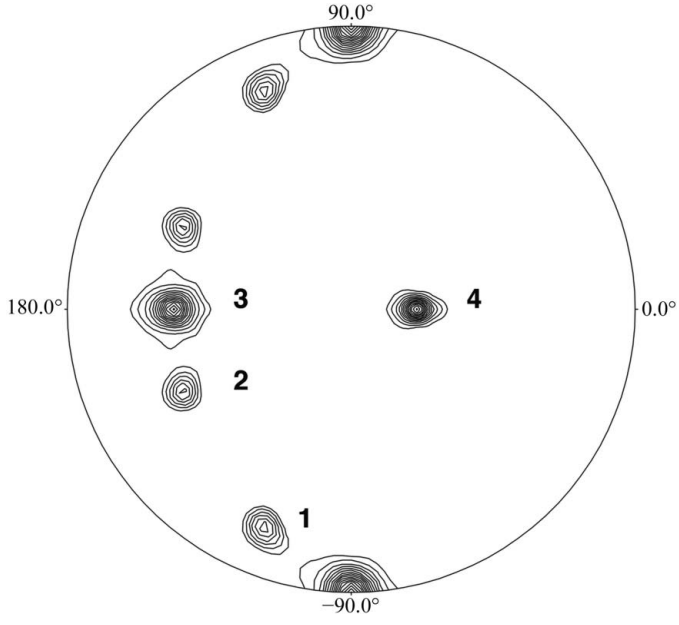

Figure 3
Self-rotation function, κ = 180° section, integration radius 20 Å, calculated at 41.5–2 Å resolution. The peak levels start at 1 r.m.s. with increments of 0.5 r.m.s. Four independent noncrystallographic twofold-axis peaks are observed and are numbered 1 to 4.

phase modification were obtained using *SHELXE*. After density modification by *DM* (Cowtan, 1994), auto model building was implemented in *ARP*/*wARP* (Perrakis *et al.*, 1999). At the zeroth cycle of autobuilding, 72% of the amino-acid residues were docked into the electron-density map. After 50 cycles, 90% of the structure was constructed with side chains, revealing the presence of four independent molecules in the asymmetric unit. This finding complies with the four peaks found in the self-rotation function. Model building of the remaining part of the structure and refinement are currently in progress.

OA gratefully acknowledges the organizers and tutors of the EMBO'07 course on 'Exploiting Anomalous Scattering in Macromolecular Structure Determination'. Data collection and processing were conducted as a part of the course. Specifically, we would like to thank Drs Wolfgang Kabsch, Kay Diederichs, Zbigniew Dauter, Santosh Panjikar, Andrew McCarthy and Sasha Popov for their guidance during this course. We would also like to thank the ESRF for synchrotron beam time and the staff scientists of the ID23-1 station for their assistance. This research was supported by the Israel Science Foundation (Grant Nos. 422/05 and 159/07) and by grants from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel.

References

Adams, J. J., Pal, G., Jia, Z. & Smith, S. P. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 305–310.
Bayer, E. A., Belaich, J. P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.
Bayer, E. A., Morag, E. & Lamed, R. (1994). *Trends Biotechnol.* **12**, 379–386.
Beguin, P., Millet, J., Chauvaux, S., Salamitou, S., Tokatlidis, K., Navas, J., Fujino, T., Lemaire, M., Raynaud, O., Daniel, M.-K. & Aubert, J.-P. (1992). *Biochem. Soc. Trans.* **20**, 42–46.
Carvalho, A. L., Dias, F. M., Nagy, T., Prates, J. A., Proctor, M. R., Smith, N., Bayer, E. A., Davies, G. J., Ferreira, L. M., Romao, M. J., Fontes, C. M. & Gilbert, H. J. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 3089–3094.
Carvalho, A. L., Dias, F. M., Prates, J. A., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M., Romao, M. J. & Fontes, C. M. (2003). *Proc. Natl. Acad. Sci. USA*, **100**, 13809–13814.
Carvalho, A. L., Pires, V. M., Gloster, T. M., Turkenburg, J. P., Prates, J. A., Ferreira, L. M., Romao, M. J., Davies, G. J., Fontes, C. M. & Gilbert, H. J. (2005). *J. Mol. Biol.* **349**, 909–915.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Ding, S. Y., Rincon, M. T., Lamed, R., Martin, J. C., McCrae, S. I., Aurilia, V., Shoham, Y., Bayer, E. A. & Flint, H. J. (2001). *J. Bacteriol.* **183**, 1945–1953.
Flint, H. J. (1997). *Trends Microbiol.* **5**, 483–488.
Jindou, S., Borovok, I., Rincon, M. T., Flint, H. J., Antonopoulos, D. A., Berg, M. E., White, B. A., Bayer, E. A. & Lamed, R. (2006). *J. Bacteriol.* **188**, 7971–7976.
Julliand, V., de Vaux, A., Millet, L. & Fonty, G. (1999). *Appl. Environ. Microbiol.* **65**, 3738–3741.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Krause, D. O., Dalrymple, B. P., Smith, W. J., Mackie, R. I. & McSweeney, C. S. (1999). *Microbiology*, **145**, 1797–1807.
Leibovitz, E. & Beguin, P. (1996). *J. Bacteriol.* **178**, 3077–3084.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Nelson, K. E., Zinder, S. H., Hance, I., Burr, P., Odongo, D., Wasawo, D., Odenyo, A. & Bishop, R. (2003). *Environ. Microbiol.* **5**, 1212–1220.
Noach, I., Frolow, F., Jakoby, H., Rosenheck, S., Shimon, L. J. W., Lamed, R. & Bayer, E. A. (2005). *J. Mol. Biol.* **348**, 1–12.
Noach, I., Lamed, R., Xu, Q., Rosenheck, S., Shimon, L. J. W., Bayer, E. A. & Frolow, F. (2003). *Acta Cryst.* D**59**, 1670–1673.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Rincon, M. T., Cepeljnik, T., Martin, J. C., Barak, Y., Lamed, R., Bayer, E. A. & Flint, H. J. (2007). *J. Bacteriol.* **189**, 4774–4783.
Rincon, M. T., Cepeljnik, T., Martin, J. C., Lamed, R., Barak, Y., Bayer, E. A. & Flint, H. J. (2005). *J. Bacteriol.* **187**, 7569–7578.
Rincon, M. T., Ding, S. Y., McCrae, S. I., Martin, J. C., Aurilia, V., Lamed, R., Shoham, Y., Bayer, E. A. & Flint, H. J. (2003). *J. Bacteriol.* **185**, 703–713.
Rincon, M. T., Martin, J. C., Aurilia, V., McCrae, S. I., Rucklidge, G. J., Reid, M. D., Bayer, E. A., Lamed, R. & Flint, H. J. (2004). *J. Bacteriol.* **186**, 2576–2585.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Shimon, L. J., Bayer, E. A., Morag, E., Lamed, R., Yaron, S., Shoham, Y. & Frolow, F. (1997). *Structure*, **5**, 381–390.
Shoham, Y., Lamed, R. & Bayer, E. A. (1999). *Trends Microbiol.* **7**, 275–281.
Spinelli, S., Fierobe, H. P., Belaich, A., Belaich, J. P., Henrissat, B. & Cambillau, C. (2000). *J. Mol. Biol.* **304**, 189–200.
Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination. A Practical Guide*, p. 402. London: Macmillan.
Tavares, G. A., Beguin, P. & Alzari, P. M. (1997). *J. Mol. Biol.* **273**, 701–713.
Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
Wedekind, K. J., Mansfield, H. R. & Montgomery, L. (1988). *Appl. Environ. Microbiol.* **54**, 1530–1535.