Structural characterization of a novel autonomous cohesin from *Ruminococcus flavefaciens*Milana Voronov-Goldman,^{a,b,‡}
Maly Levy-Assaraf,^{a,b,‡} Oren
Yaniv,^{a,b} Gloria Wisserman,^a
Sadanari Jindou,^{a,c} Ilya
Borovok,^a Edward A. Bayer,^d
Raphael Lamed,^{a,b} Linda J. W.
Shimon^{e,*} and Felix Frolow^{a,b,*}^aDepartment of Molecular Microbiology and
Biotechnology, Tel Aviv University, Tel Aviv
69978, Israel, ^bThe Daniella Rich Institute for
Structural Biology, Tel Aviv University, Tel Aviv
69978, Israel, ^cFaculty of Agriculture, Meijo
University, Nagoya 468-8502, Japan,
^dDepartment of Biological Chemistry, The
Weizmann Institute of Science, Rehovot 76100,
Israel, and ^eDepartment of Chemical Research
Support, The Weizmann Institute of Science,
Rehovot 76100, Israel‡ These authors contributed equally to this
work.Correspondence e-mail:
linda.shimon@weizmann.ac.il,
mbfrolow@post.tau.ac.ilReceived 16 January 2014
Accepted 20 February 2014PDB reference: cohesin from *R. flavefaciens*,
4n2o © 2014 International Union of Crystallography
All rights reserved

Ruminococcus flavefaciens is a cellulolytic bacterium found in the rumen of herbivores and produces one of the most elaborate and variable cellulosome systems. The structure of an *R. flavefaciens* protein (RfCohG, ZP_06142108), representing a freestanding (non-cellulosomal) type III cohesin module, has been determined. A selenomethionine derivative with a C-terminal histidine tag was crystallized and diffraction data were measured to 2.44 Å resolution. Its structure was determined by single-wavelength anomalous dispersion, revealing eight molecules in the asymmetric unit. RfCohG exhibits the most complex among all known cohesin structures, possessing four α -helical elements and a topographical protuberance on the putative dockerin-binding surface.

1. Introduction

Cellulosomes are extracellular multi-enzyme complexes designed for efficient degradation of cellulose and related plant cell-wall polysaccharides (Lamed *et al.*, 1983). The cellulosomal enzymes are assembled on a noncatalytic scaffoldin protein comprised of an array of repeating cohesin modules. The enzymes carry a dockerin module, which interacts with a single cohesin of a specific scaffoldin. The selective and high-affinity binding of the dockerin-bearing enzyme subunits to the scaffoldin-borne cohesins dictates the assembly of the mature cellulosome. The integrity of the entire cellulosome complex is thus maintained by the cohesin–dockerin interaction (Xu *et al.*, 2004; Bayer *et al.*, 1994, 1998).

Ruminococcus flavefaciens is a predominant cellulolytic rumen bacterium, which forms a multi-enzyme cellulosome complex that plays a central role in the ability of this bacterium to degrade plant cell-wall polysaccharides (Antonopoulos *et al.*, 2004; Jullian *et al.*, 1999; Krause *et al.*, 1999; Sijpesteijn, 1951). *R. flavefaciens* strain FD-1 possesses the most complex cellulosome system known to date (Fig. 1), as indicated by its genome sequence (Berg Miller *et al.*, 2009). Nearly 225 different dockerin-containing coding sequences have been identified in the *R. flavefaciens* FD-1 genome (Rincon *et al.*, 2010). Surprisingly, a large number of these ORFs do not encode the types of protein that one would expect for cellulosomes, *i.e.* carbohydrate-active enzymes, but rather novel and unexpected structural and catalytic modules annotated as putative proteases, trans-glutaminases, LRR proteins, serpins and many small novel scaffoldin sequences (composed of a dockerin and one or two suspected cohesin modules). The role of these proteins has yet to be correlated with the established cellulosome function of enhancing polysaccharide degradation.

Interestingly, several key cellulosomal scaffoldins, including ScaA, ScaB, ScaC, CttA and ScaE, are organized into a single *sca* gene cluster. The presence of this gene cluster has been documented in at least five different strains of *R. flavefaciens* (Jindou *et al.*, 2008). Additionally, one peculiar and very interesting ORF containing only a signal peptide and an autonomous cohesin module (*i.e.* technically not part of a scaffoldin) was identified by bioinformatics analysis and termed RfCohG. Biochemical evidence including enzyme-linked immunosorbent assays (ELISA) has shown that the RfCohG binds

cellulosome-related proteins such as ScaB and CttA (unpublished data). To understand the function and framework of this intriguing cellulosome-free *RfCohG* and its relation to cellulosomal cohesins, we have undertaken the determination of its three-dimensional structure.

2. Materials and methods

2.1. Cloning

The DNA encoding for the *RfCohG* module (accession No. ZP_06142108, residues 26–218) was amplified by PCR from *R. flavefaciens* FD-1 genomic DNA, isolated as described by Murray & Thompson (1980), using the primers 5'-CATGCCATGGGAG-CAGTTCGGTTACTGCTGATCTG and 5'-CCGCTCGAGTTC-AACTGTTATAGTGCCGCCCTCC designed according to the DNA sequence of the genomic Contig43 (GenBank NZ_ACOK01000043). The PCR product was inserted into pET28a(+) expression vector (Novagen, Madison, Wisconsin, USA) via *Nco*I and *Xho*I restriction enzymes to introduce a C-terminal hexahistidine (His) tag.

2.2. Expression of seleno-L-methionine-labelled protein

Expression was conducted according to the method described earlier (Van Duyne *et al.*, 1993), with minor modifications. Transformed cells from a culture grown overnight in 1 ml Luria–Bertani

broth containing 50 $\mu\text{g ml}^{-1}$ kanamycin were isolated and resuspended in 1 ml M9 minimal medium (Sigma, St Louis, Missouri, USA), supplemented with glucose (4 mg ml^{-1}), 1 ml 100 mM CaCl_2 , 1 ml 1 M MgSO_4 , thiamine and vitamin B₁ (5 mg ml^{-1} each). The resuspended culture was 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 OD₆₀₀ of 0.6. At this point, seleno-L-methionine was added to a final concentration of 50 $\mu\text{g ml}^{-1}$. Inhibition of the methionine pathway was achieved with the addition of 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 mM isopropyl β -D-1-thiogalactopyranoside was added and the culture was grown for an additional 13 h.

Cells were harvested by centrifugation (4000g for 15 min) at 277 K and resuspended in 50 mM NaH_2PO_4 pH 8.0 containing 300 mM NaCl at a ratio of 1 g wet pellet to 4 ml buffer solution. DNase was added prior to the sonication procedure. The suspended pellet was sonicated in a sonicator Ultrasonic Processor X2 (Misonix Inc.) for 20 min in discontinuous mode (0.05 s pulse on and 0.05 s pulse off). The suspension was kept on ice during sonication; it was then centrifuged (20 000g at 277 K for 20 min) and the supernatant fluids were collected.

Figure 1

Schematic overview of the modular interactions in the cellulosome system of *R. flavefaciens* strain FD-1. The primary two-cohesin scaffoldin ScaA binds specifically either to the Cel44A-type dockerins (a) or to the C-terminal dockerin of ScaC (b). The ScaA dockerin module interacts with one of the five type III cohesins (numbered 5–9) (c) carried by the adaptor scaffoldin ScaB. Like the ScaA cohesins, ScaB cohesins 1–4 (d) also bind both the ScaC dockerin (b) and Cel44A-type dockerins (a). The conserved XDoc dyad (in red) of ScaB (e) interacts with the cohesin module of the anchoring scaffoldin ScaE. The ScaE cohesin also interacts with the XDoc dyads of the cellulose-binding proteins CttA and the putative cysteine peptidase RflaF_05439 (Levy-Assaraf *et al.*, 2013) (f). An additional standalone cohesin, CohG, the topic of this communication, interacts selectively with the XDoc dyads of ScaB and CttA but not with that of RflaF_05439.

The recombinant His-tagged protein was first isolated by metal-chelate affinity chromatography using Ni-IDA resin (Rimon Biotech, Israel) according to the manufacturer's recommended protocol. Further purification was accomplished by fast protein liquid chromatography using a Superdex 75 16/60 column and an ÄKTAprime system (GE Healthcare). The purified protein solution consisted of

Screening numerous SeMet-*RfCohG* crystals finally yielded a crystal that gave a reasonable diffraction pattern extending to a resolution of 2.44 Å with well separated diffraction spots. The data set was collected in 1° oscillations, and a total of 360 images were collected, indexed, processed and scaled using *DENZO* and *SCALEPACK* as implemented in *HKL-2000* (Otwinowski & Minor, 1997). The crystal belonged to the monoclinic space group *C2*. The calculated Matthews coefficient (Matthews, 1968) was 3.18 Å³ Da⁻¹, corresponding to the presence of approximately eight molecules in the asymmetric unit. The statistics of the diffraction data are shown in Table 1.

Table 1

Crystal parameters and data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

X-ray source	ESRF ID29
Space group	C2
No. of crystals	1
Total rotation angle (°)	360
Unit-cell parameters	
<i>a</i> (Å)	116.24
<i>b</i> (Å)	67.73
<i>c</i> (Å)	258.91
β (°)	93.08
<i>V</i> (Å ³)	2038247
No. of molecules in asymmetric unit	8
Resolution range (Å)	50–2.44 (2.47–2.44)
Total No. of reflections	554681
Unique reflections	74822 (3735)
Mosaicity (°)	0.48–1.14
Multiplicity	7.34 (6.0)
Completeness (%)	98.8 (97.3)
Mean <i>I</i> / σ (<i>I</i>)	14.4 (2.0)
<i>R</i> _{merge} †	0.135
Overall average <i>B</i> factor (Å ²)	45.8
No. of protein residues	1544
No. of solvent atoms	167
No. of ions	1
<i>R</i> _{cryst} / <i>R</i> _{free}	0.20542/0.22941
Geometry	
R.m.s.d., bonds (Å)	0.011
R.m.s.d., angles (°)	1.976
MolProbity validation	
Ramachandran favoured (%) (goal >98%)	99.33
Ramachandran outliers (%) (goal <0.2%)	0.0
Rotamer outliers (goal <1%)	1.5
C β outliers (goal 0)	0
Clashscore‡	4.43

† $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\sum_{hkl}$ denotes the sum over all reflections and $\sum_i$ the sum over all equivalent and symmetry-related reflections (Stout & Jensen, 1968). ‡ Clashscore is the number of serious steric overlaps (>0.4 Å) per 1000 atoms.

2.5. Structure determination and refinement

The structure of SeMet-*RfCohG* was determined by the single-wavelength anomalous dispersion (SAD) method. The programs *SHELXC/D/E* (Sheldrick, 2010) as implemented in *HKL2MAP* (Pape & Schneider, 2004) were used in heavy-atom substructure determination, phase determination and polyaniline tracing. The default decision of the programs regarding the starting parameters to be used in substructure determination proved unsuccessful. Manual forcing of the complete resolution and increasing the number of trials to 1000 led to the determination of the substructure. Subsequently, the protein structure was phased, phases were modified, and the initial polyaniline skeleton was built, exhibiting eight independent molecules in the asymmetric unit. *ARP/wARP* was used for building the structure (Langer *et al.*, 2008). *REFMAC5* (Murshudov *et al.*, 2011) and *Coot* (Emsley *et al.*, 2010) were used for structure refinement and rebuilding. The final coordinates and structure factors were deposited in the PDB under accession code 4n2o.

3. Results and discussion

3.1. Overall structure of *RfCohG* reveals jelly-roll topology decorated by four α -helical elements

RfCohG forms an elongated nine-stranded β -sandwich in a classical jelly-roll topology with approximate overall dimensions of 51 × 35 × 28 Å as measured on the C α skeleton using *PyMOL* (DeLano, 2002). The 'front face' of the molecule is formed by β -strands 8, 3, 6, 5 and α -helices 1, 2 and 4, and the 'back face' is formed by β -strands 9, 1, 2, 7, 4 and α -helix 3. As shown in Fig. 2, strands 1

and 9 are aligned parallel to each other, while the other β -strands are anti-parallel. β -Strands 4 and 8 are disrupted by β -flaps (Fig. 2, coloured green). There are four distinct α -helices in *RfCohG*. These include a seven-residue α -helix (H1) in the N-terminal region, a three-residue α -helix (H2) located on the β -flap between β -strands 3 and 4, a five-residue α -helix (H3) located between β -strands 5 and 6, and a ten-residue α -helix (H4) at the C-terminus (see Figs. 2 and 3). There are eight molecules per asymmetric unit, organized as two tetramers, each of which is a dimer of dimers (Fig. 4a). These independent molecules are very similar with r.m.s.d.s of between 0.03 and 0.13 Å on C α positions. Upon operation of symmetry elements of space group C2, the structure packs into helical fibre-like formations (Fig. 4b).

3.2. Comparison of *RfCohG* with other cohesin structures

A structural similarity search, using the *PDBFold* server (European Bioinformatics Institute; <http://www.ebi.ac.uk/msd-srv/ssm>;

Figure 4

Stereoview (cross-eye) of (a) eight independent molecules of the asymmetric unit of *RfCohG* and (b) two adjacent fibre-like formations formed by applying the symmetry elements of space group C2 to the asymmetric unit. Note that head-to-head contacts (A–A and G–G) between asymmetric units are formed by twofold axes.

Figure 5

Structural comparison among different cellulosomal cohesin types. (a) Type I cohesin (PDB entry 1anu), (b) type II cohesin (PDB entry 1tyj), (c) the type IIIe cohesin *RfCohE* (PDB entry 2zf9), (d) the type III cohesin *RfCohG*. α -Helical elements in each molecule are coloured yellow, β -strands are coloured blue and β -flaps are coloured green.

Figure 6

Structural comparison of *RfCohG* with *RfCohE*. (a) Cartoon representation of the superposition of $C\alpha$ atoms of *RfCohG* (blue) and *RfCohE* (red). (b) Surface and cartoon representations highlighting the protrusion of *RfCohG* (coloured cyan), composed of β -flaps 4 and 8 and α -helix 2, and the protrusion of *RfCohE* (coloured magenta) composed of β -flap 8. (c) Molecules in (b) rotated 90° about the y axis, emphasizing the height of the protrusion.

Krissinel & Henrick, 2004) revealed that despite the relatively low sequence identity (9–25%), *RfCohG* can be structurally aligned to cohesin modules from all of the known types (I–III). *RfCohG* is the most similar to the type IIIe cohesin from *R. flavefaciens* ScaE (PDB entry 2zf9; Alber *et al.*, 2009; Salama-Alber *et al.*, 2013) with an r.m.s.d. of 1.36 Å, compared with an r.m.s.d. of 2.29 Å with the type I cohesin from the scaffoldin subunit of *Clostridium thermocellum* (PDB entry 1anu; Shimon *et al.*, 1997) and an r.m.s.d. of 2.8 Å with the type II cohesin from the scaffoldin subunit of *Bacteroides cellulosolvens* (PDB entry 1tyj; Noach *et al.*, 2005). The positions of the

RfCohG β -flaps are consistent with those of the type II cohesins, *i.e.* in the midst of β -strands 4 and 8 (Noach *et al.*, 2003, 2005; Carvalho *et al.*, 2005).

The addition of the secondary-structural elements is clearly observed upon comparison between the different cohesin types (Figs. 3 and 5). In this context, type I cohesins are constrained to form only β -sheets or extremely short α -helix motifs (Fig. 5a), type I cohesins possess longer α -helices (Fig. 5b) in addition to two β -flaps, and type IIIe cohesins possess one or more dominant α -helix elements in addition to these two β -flaps (Figs. 5c and 5d).

Moreover, *RfCohG* possesses a unique structural feature that distinguishes it from the other known cohesin structures. While the type II and type III cohesins (*RfCohE* from *ScaE*) possess a single α -helix located at different positions in the cohesin molecule (Noach *et al.*, 2005; Alber *et al.*, 2009), *RfCohG* possesses four α -helical elements (indicated in yellow in Fig. 5). The position of the last helix (H4) is equivalent to the position of the dominant helix in the *RfCohE* type III molecule (Fig. 3), although the helix of *RfCohG* is four residues shorter. All four α -helices are separated from the major aligned portion of the molecule. α -Helix 2 (marked H2 in Fig. 2) and the short β -strand 4 are both located within the lengthy β -flap connecting strands 3 and 5. This β -flap protrudes 11 Å from the planar face of the 8–3–6–5 β -sheet, which together with β -flap 4 forms the protuberance in the middle of the molecule (Fig. 6).

3.3. Three tandem tyrosine residues anchor the helices to the molecule core

Similar to other cohesin modules (Shimon *et al.*, 1997; Alber *et al.*, 2009), the nine β -strands of *RfCohG* are assembled around a hydrophobic core of aromatic and aliphatic residues with numerous stabilizing aromatic interactions (Burley & Petsko, 1985).

Interestingly, the core includes a distinctive stretch of three repeated tyrosines (Tyr160, Tyr161 and Tyr162), located in the linker region between β -strand 8a and β -strand 8b (Fig. 3, highlighted in cyan, Fig. 7). Tyr 162 connects α -helices 1 and 3 *via* hydrophobic interactions with Val14, Ala23, Phe24 and Phe182. Tyr161 connects α -helix 2 by hydrophobic interactions with Ile15, Ile71, Trp107, Pro112, Val114 and Leu115. Tyr 160 consolidates the α -helices to the rest of the molecule *via* hydrophobic interactions with Ala13, Val14, Ile37, Leu39, Ile68 and Ile70. This composite feature may serve as an example whereby an extensive hydrophobic core in a molecule connects and melds distinct secondary-structural elements together. In the related *RfScaE* structure only one of these three tyrosines is conserved but the adjacent β -sheets are close enough (5.2 Å) to be stabilized by main-chain hydrogen bonding.

Two hydrophobic clusters are observed in the structure of the *RfCohG* molecule. The first large hydrophobic cluster at one end of the cohesin comprises aromatic residues (Phe24, Phe129, Phe168,

Tyr177, Tyr180 and Phe182) and aliphatic residues (Val27, Ile37 and Val56). The second hydrophobic cluster at the centre of the molecule includes aromatic residues (Tyr72, Phe90, Trp107, Phe124, Phe142 and Phe144), connecting the front and the back of the molecule and the loops, together with numerous aliphatic residues (Val14, Ile15, Leu39, Ile68, Ile70, Ile71, Ile72, Val88, Leu96, Ile102, Leu111, Val114, Leu123, Leu125, Leu158 and Ile190).

An additional observation that supports the stabilizing role of the α -helices in the *RfCohG* structure is evident from interface calculations between the two β -sheets of the molecule. Using the PISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html; Krissinel & Henrick, 2007), analysis of the hydrophobic contacts in the core of the *RfCohG* and *RfCohE* structures was performed by calculating the contact area between the two β -sheets that form the β -sandwich with the α -helices included and between the same β -sheets only (eliminating the other secondary structures). Using this approach, it was found that *RfCohG* possesses a larger interface area of 912 Å² compared with 802 Å² for *RfCohE*. Moreover, when the calculation was performed without helix regions, the interface was found to be 412 and 681 Å² for *RfCohG* and *RfCohE*, respectively. The overall dimensions of β -sheet regions from *RfCohG* and *RfCohE* are very similar (r.m.s.d. of 2.8 Å). It thus follows that α -helical regions contribute to more than half of the interface area in the *RfCohG* molecule.

The additional helical elements revealed in the *RfCohG* structure show a continuum of progressive cohesin complexity (Fig. 5). The helices and the hydrophobic core may impart robustness to *RfCohG* and may play a role in the as yet unknown function of this unusual, freestanding cohesin.

We thank the ESRF, Grenoble, France, for use of the macromolecular crystallographic data-collection facilities and the ID29 staff for their assistance. This research was supported by the Israel Science Foundation (ISF, grant Nos. 715/12 and 1349/13) and by the Sidney E. Frank Foundation through the Israel Science Foundation (grant No. 24/11). Additional support was received from the US-Israel Binational Science Foundation (BSF) and Israeli Center of Research Excellence (I-CORE Center No. 152/11) managed by the Israel Science Foundation. EAB is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-Organic Chemistry at the Weizmann Institute of Science.

References

- Alber, O., Noach, I., Rincon, M. T., Flint, H. J., Shimon, L. J. W., Lamed, R., Frolov, F. & Bayer, E. A. (2009). *Proteins*, **77**, 699–709.
- Antonopoulos, D. A., Nelson, K. E., Morrison, M. & White, B. A. (2004). *Environ. Microbiol.* **6**, 335–346.
- Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. (1998). *Curr. Opin. Struct. Biol.* **8**, 548–557.
- Bayer, E. A., Morag, E. & Lamed, R. (1994). *Trends Biotechnol.* **12**, 379–386.
- Berg Miller, M. E. *et al.* (2009). *PLoS One*, **4**, e6650.
- Burley, S. K. & Petsko, G. A. (1985). *Science*, **229**, 23–28.
- Carvalho, A. L., Pires, V. M. R., Gloster, T. M., Turkenburg, J. P., Prates, J. A. M., Ferreira, L. M. A., Romão, M. J., Davies, G. J., Fontes, C. M. G. A. & Gilbert, H. J. (2005). *J. Mol. Biol.* **349**, 909–915.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486–501.
- Jindou, S., Brulc, J. M., Levy-Assaraf, M., Rincon, M. T., Flint, H. J., Berg, M. E., Wilson, M. K., White, B. A., Bayer, E. A., Lamed, R. & Borovok, I. (2008). *FEMS Microbiol. Lett.* **285**, 188–194.
- Julliard, V., de Vaux, A., Millet, L. & Fonty, G. (1999). *Appl. Environ. Microbiol.* **65**, 3738–3741.
- Krause, D. O., Dalrymple, B. P., Smith, W. J., Mackie, R. I. & McSweeney, C. S. (1999). *Microbiology*, **145**, 1797–1807.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst. D* **60**, 2256–2268.

Figure 7
Stabilization of the spatial position of the three α -helices by packing around three tandem tyrosine residues. The three tandem tyrosine residues and their hydrophobically interacting partners are presented as sticks coloured as follows: Tyr160 in green, Tyr161 in red and Tyr162 in purple.

- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Lamed, R., Setter, E. & Bayer, E. A. (1983). *J. Bacteriol.* **156**, 828–836.
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
- Levy-Assaraf, M., Voronov-Goldman, M., Grinberg, I. R., Weiserman, G., Shimon, L. J. W., Jindou, S., Borovok, I., White, B. A., Bayer, E. A., Lamed, R. & Frolov, F. (2013). *PLoS One*, **8**, e56138.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*, 1st ed. New York: Wiley.
- Murray, M. G. & Thompson, W. F. (1980). *Nucleic Acids Res.* **8**, 4321–4325.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Noach, I., Frolov, 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. & Frolov, F. (2003). *Acta Cryst.* **D59**, 1670–1673.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
- Rincon, M. T., Dassa, B., Flint, H. J., Travis, A. J., Jindou, S., Borovok, I., Lamed, R., Bayer, E. A., Henrissat, B., Coutinho, P. M., Antonopoulos, D. A., Berg Miller, M. E. & White, B. A. (2010). *PLoS One*, **5**, e12476.
- Salama-Alber, O., Jobby, M. K., Chitayat, S., Smith, S. P., White, B. A., Shimon, L. J. W., Lamed, R., Frolov, F. & Bayer, E. A. (2013). *J. Biol. Chem.* **288**, 16827–16838.
- Sheldrick, G. M. (2010). *Acta Cryst.* **D66**, 479–485.
- Shimon, L. J., Bayer, E. A., Morag, E., Lamed, R., Yaron, S., Shoham, Y. & Frolov, F. (1997). *Structure*, **5**, 381–390.
- Sijpesteijn, A. K. (1951). *J. Gen. Microbiol.* **5**, 869–879.
- Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination. A Practical Guide*. London: McMillan.
- Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
- Xu, Q., Bayer, E. A., Goldman, M., Kenig, R., Shoham, Y. & Lamed, R. (2004). *J. Bacteriol.* **186**, 968–977.