Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure and function of the *Clostridium thermocellum* cellobiohydrolase A X1-module repeat: enhancement through stabilization of the CbhA complex

Roman Brunecky, Markus Alahuhta, Yannick J. Bomble, Qi Xu, John O. Baker, Shi-You Ding, Michael E. Himmel and Vladimir V. Lunin*

Biosciences Center, National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, CO 80401, USA

Correspondence e-mail: vladimir.lunin@nrel.gov

Received 10 November 2011
Accepted 13 January 2012

PDB References: $X1_1$, 3pdg; $X1_2$, 3pe9; $X1_1$-$X1_1$, 3pdd.

The efficient deconstruction of lignocellulosic biomass remains a significant barrier to the commercialization of biofuels. Whereas most commercial plant cell-wall-degrading enzyme preparations used today are derived from fungi, the cellulosomal enzyme system from *Clostridium thermocellum* is an equally effective catalyst, yet of considerably different structure. A key difference between fungal enzyme systems and cellulosomal enzyme systems is that cellulosomal enzyme systems utilize self-assembled scaffolded multimodule enzymes to deconstruct biomass. Here, the possible function of the X1 modules in the complex multimodular enzyme system cellobiohydrolase A (CbhA) from *C. thermocellum* is explored. The crystal structures of the two X1 modules from *C. thermocellum* CbhA have been solved individually and together as one construct. The role that calcium may play in the stability of the X1 modules has also been investigated, as well as the possibility that they interact with each other. Furthermore, the results show that whereas the X1 modules do not seem to act as cellulose disruptors, they do aid in the thermostability of the CbhA complex, effectively allowing it to deconstruct cellulose at a higher temperature.

1. Introduction

Cellulose is digested naturally by microorganisms using different combinations of enzymes and accessory proteins. Various strategies have evolved to address the structurally and chemically diverse nature of the cellulosic biomass. This natural diversity of cell-wall digestion schemes provides an opportunity to improve current catalyst formulations by the incorporation of critical components with unique properties dictated by the substrate. However, the functions of many components of these complex processes are not well understood.

The aggregation of glycoside hydrolases into cellulosomal structures is a strategy used by some bacteria and a few anaerobic fungi (Bayer *et al.*, 2004). Some organisms utilize noncomplexed as well as cellulosomal complexes to hydrolyze plant cell walls (Bayer *et al.*, 2004). Cellulosomes consist of protein scaffolds containing cohesin modules which bind to dockerin modules linked to glycoside hydrolases and accessory protein modules with different functionalities. These complex proteins work together as a complex tethered to the surface of the bacterial cell. CbhA is a large multimodular enzyme found to be prevalent in *Clostridium thermocellum* cellulosomes (Gold & Martin, 2007). CbhA (Zverlov *et al.*, 1998) has seven modules; from the N-terminus, these are CBM4 (Alahuhta *et al.*, 2010), an immunoglobulin-like module (Kataeva *et al.*, 2004), GH9 (Schubot *et al.*, 2004), two

© 2012 International Union of Crystallography
Printed in Singapore – all rights reserved

X1 modules (Kataeva *et al.*, 2002, 2003), CBM3b (Jindou *et al.*, 2007) and finally a type I dockerin module.

Bacterial X1 modules (once called fibronectin III-like modules) are commonly found in nature (Little *et al.*, 1994). Despite being structurally very similar to fibronectin III-like modules, these CbhA modules were reclassified in 2004 by Devillard and coworkers as X modules (Devillard *et al.*, 2004) and have more recently been referred to as X1 modules (Kataeva *et al.*, 2005), a nomenclature that we have adopted. They can be expressed separately from the cellulosomal multimodule enzymes as independently folded proteins. Some studies have proposed that X1 modules may be cellulose disruptors that improve the hydrolytic ability of cellulases (Kataeva *et al.*, 2002) or affect the processivity of the cellulase module (Chiriac *et al.*, 2010). There is also convincing evidence that these structures are simply a 'storage form' of the linker peptide that connects other modules of the enzyme that can be extended as needed. Alternatively, they may simply function as spacers between the other modules (Watanabe *et al.*, 1994; Jee *et al.*, 2002; Adams *et al.*, 2010). Unfolding studies of CbhA show that multimodule constructs show more cooperative unfolding than individual modules, suggesting that module coupling by X1 modules is a significant stabilizing factor (Kataeva *et al.*, 2005). Additionally, work by Adams and coworkers also suggests that X modules can directly interact with the cohesin $CohI_9$, forming substantial van der Waals contacts (Adams *et al.*, 2010).

It is our long-term strategy to study and understand the structure–function relationships governing the action of this large cellulosomal enzyme complex. The crystal structures of the CbhA Ig-GH9 construct (Schubot *et al.*, 2004), CBM3b (Jindou *et al.*, 2007) and the CBM4-Ig construct (Alahuhta *et al.*, 2010) have recently been solved and form the basis of our current structural understanding. To complete the structural ensemble, we present here the X-ray diffraction structures of the two X1 modules together with computer-simulation and biochemical characterization results.

2. Materials and methods

2.1. Expression, purification and crystallization

Three fragments consisting of X1 modules of the CbhA gene were amplified by PCR: the first module $X1_1$, the second module $X1_2$ and both modules together in the same construct $X1_1$-$X1_2$. The following primers were used: AAACCG**CCATGG**TAACTATTGATTCGCCTG and TACAGT**CTCGAG**CACAATTACCTTAACAAGTA for $X1_1$, TGAAAC**CCATGG**TAAAACTTACTGCACCCAA and TACTTT**CTCGAG**CCGTGCCTGTTTTACAAATA for $X1_2$, AAACGC**CCATGG**TAACTATTGATTCGCCTG and TACTTT**CTCGAG**CCGTGCCTGTTTTACAAATA for $X1_1$-$X1_2$ (the bold oligonucleotides indicate restriction sites). The template used for all PCR was *C. thermocellum* genomic DNA. The resultant PCR fragments were inserted into the plasmid pET28a (Novagen, Madison, Wisconsin, USA) *via* the *Nco*I and *Xho*I restriction sites to generate the respective expression plasmids. The two truncated CbhAs, CBM4-Ig-GH9-$X1_1$-$X1_2$ and CBM4-Ig-GH9, were amplified by PCR using the primers TCCGTGCATATGTTAGAAGATAATTCTTCGACT, AATAGTCTCGAGATCGGTTTCACTGTCTGTGT and CTGTACCTCGAGATCCCGTGCCTGTTTTACAA as described above. These two PCR fragments were inserted into plasmid pET22b (Novagen, Madison, Wisconsin, USA) *via* the *Nde*I and *Xho*I restriction sites to generate the respective expression plasmids. All plasmids inserted with target genes were transformed into *Escherichia coli* (BL21) (Agilent, Santa Clara, California, USA) and overexpression was performed at 289 or 310 K in the presence of 0.3 m*M* IPTG. All recombinant proteins contained a C-terminal His tag (six histidines) and were purified using the QIAexpress Ni–NTA protein-purification system (Qiagene, Valencia, California, USA).

$X1_1$ was additionally purified by cation-exchange chromatography using a Source 15S column (GE Healthcare, Piscataway, New Jersey, USA). For this chromatography, buffer *A* was 20 m*M* acetic acid pH 5.5 and buffer *B* was 20 m*M* acetic acid pH 5.5 with 1 *M* NaCl. $X1_2$ was purified using a Source 15PHE hydrophobic interaction column (GE Healthcare, Piscataway, New Jersey, USA) with buffer *A* and buffer *C* (20 m*M* acetate pH 5.0 with 1 *M* ammonium sulfate). Relevant fractions were then subjected to cation-exchange chromatography using a Source 15S column (GE Healthcare, Piscataway, New Jersey, USA) and eluted with buffer *D* (20 m*M* Tris–HCl pH 6.8) and buffer *E* (20 m*M* Tris–HCl pH 6.8 with 1 *M* NaCl). $X1_1$-$X1_2$ was further purified by anion-exchange chromatography using a Source 15Q column (GE Healthcare, Piscataway, New Jersey, USA) and buffer *F* (20 m*M* Tris–HCl pH 8.0) and buffer *G* (20 m*M* Tris–HCl pH 8.0 with 1 *M* NaCl). The CBM4-Ig-GH9-$X1_1$-$X1_2$ construct was purified by anion-exchange chromatography using a Source 15Q column (GE Healthcare, Piscataway, New Jersey, USA) and buffer *H* (20 m*M* Tris–HCl pH 7.0) and buffer *I* (20 m*M* Tris–HCl pH 7.0 with 1 *M* NaCl). The CBM4-Ig-GH9 construct only required size-exclusion chromatography, as described below, to be free of major impurities.

Finally, all proteins were separated from minor impurities by size-exclusion chromatography using HiLoad Superdex 75 (26/60; GE Healthcare, Piscataway, New Jersey, USA) in buffer *H* (20 m*M* Tris–HCl pH 7.0) containing 100 m*M* NaCl, 1 m*M* EDTA and 1 m*M* sodium azide. The purified fusion proteins were concentrated using a Vivaspin 5K concentrator (Vivaproducts, Littleton, Massachusetts, USA) and the protein concentration was measured using a NanoDrop UV spectrophotometer (NanoDrop, Wilmington, Delaware, USA).

Diffraction-quality crystals were obtained by sitting-drop vapor diffusion using a 96-well plate. 50 µl well solution was used with drops consisting of 1 µl well solution and 1 µl protein solution. $X1_1$ solution consisted of 80 mg ml^{-1} protein, 20 m*M* Tris pH 7 and 100 m*M* NaCl. The native data set for $X1_1$ was collected from crystals grown in 1.9 *M* sodium malonate pH 6.0 using Grid Screen Sodium Malonate from Hampton Research (Aliso Viejo, California, USA). Before flash-cooling, the crystal was transferred into a 2 µl cryosolution drop consisting of 20%(*v*/*v*) ethylene glycol in the well

Table 1
X-ray data-collection and refinement statistics.

Values in parentheses are for the highest resolution bin; numbers of atoms include alternative conformations.

	$X1_1$	$X1_2$	$X1_1$-$X1_2$
Data collection			
Space group	$P4_22_12$	$P2_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 42.32, b = 42.32, c = 116.76, $\alpha = \beta = \gamma$ = 90.0	a = 42.13, b = 103.28, c = 45.63, α = 90.0, β = 106.53, γ = 90.0	a = 46.91, b = 58.09, c = 78.84, $\alpha = \beta = \gamma$ = 90.0
Wavelength (Å)	1.54178	1.54178	1.54178
Temperature (K)	100	100	100
Resolution (Å)	25.0–1.78 (1.88–1.78)	25.0–1.69 (1.78–1.69)	25.0–1.72 (1.81–1.72)
Unique reflections	10862 (1573)	41325 (5814)	23576 (3291)
Observed reflections	137621 (7094)	234312 (16511)	215248 (17244)
R_{int}†	0.061 (0.282)	0.064 (0.406)	0.111 (0.621)
Average multiplicity	12.67 (4.51)	5.67 (2.84)	9.13 (5.24)
$\langle I/\sigma(I)\rangle$	22.59 (2.72)	13.68 (2.53)	13.57 (1.96)
Completeness (%)	99.7 (98.9)	98.8 (97.6)	99.9 (100)
Refinement			
R/R_{free}	0.230 (0.308)/0.271 (0.401)	0.182 (0.282)/0.229 (0.325)	0.167 (0.318)/0.226 (0.353)
Protein atoms	717	2941	1556
Water molecules	101	606	520
Other atoms	2	17	5
R.m.s.d. from ideal bond lengths‡ (Å)	0.023	0.022	0.023
R.m.s.d. from ideal bond angles‡ (°)	1.98	1.90	1.79
Wilson B factor (Å^2)	21.2	15.7	16.3
Average B factor for protein atoms (Å^2)	14.8	8.2	9.7
Average B factor for water molecules (Å^2)	20.1	17.3	21.5
Ramachandran plot statistics§ (%)			
Allowed	100	100	100
Favored	99.1	98.5	99.1
Outliers	0	0	0

† $R_{int} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual reflection and $\langle I(hkl)\rangle$ is the mean intensity of a group of equivalents; the sums are calculated over all reflections with more than one equivalent measured. ‡ Engh & Huber (1991). § Chen *et al.* (2010).

solution. An iodine-derived data set for $X1_1$ was obtained using a crystal from the same condition (1.9 *M* sodium malonate pH 6.5). Before data collection, this sample was moved into a freshly made 2 µl drop consisting of 1.9 *M* sodium malonate pH 6.5, 20%(*v*/*v*) ethylene glycol, 50 m*M* KI (with some granular KI inside the same sealed sitting-drop well setup) and then incubated overnight. The $X1_2$ solution consisted of 15 mg ml^{-1} protein in 20 m*M* acetic acid buffer pH 5 with 100 m*M* NaCl. Crystals were grown using the PEG/Ion HT screen from Hampton Research (Aliso Viejo, California, USA) in 0.2 *M* ammonium iodide, 20%(*w*/*v*) polyethylene glycol 3350 pH 6.2. Before flash-cooling, the crystal was incubated for 5 s in a 2 µl cryosolution drop consisting of 0.2 *M* ammonium iodide, 25%(*w*/*v*) polyethylene glycol 3350 pH 6.2 with 10%(*v*/*v*) glycerol. The $X1_1$-$X1_2$ solution consisted of 21.8 mg ml^{-1} protein in 20 m*M* Tris pH 7 with 100 m*M* NaCl. Crystals were grown using the PEG/Ion HT screen from Hampton Research (Aliso Viejo, California, USA) in 0.2 *M* $CaCl_2$, 20%(*w*/*v*) PEG 3350 pH 5.1. Before data collection, the $X1_1$-$X1_2$ crystal was briefly dipped into a 2 µl cryosolution drop consisting of 10% ethylene glycol in the well solution.

2.2. X-ray diffraction and structure determination

Before data collection, all crystals were flash-cooled in a cold nitrogen-gas stream at 100 K. Data collection was performed using a Bruker X8 MicroStar X-ray generator with Helios mirrors and a Bruker PLATINUM 135 CCD detector. Data were indexed and processed with the Bruker suite of programs v.2008.10 (Bruker AXS, Madison, Wisconsin, USA). Intensities were converted into structure factors and 5% of the reflections were flagged for R_{free} calculations using the programs *F2MTZ*, *TRUNCATE*, *CAD* and *UNIQUE* from the *CCP*4 suite of programs (Winn *et al.*, 2011). The structure of $X1_1$ was solved with *SHELXC*/*D*/*E* v.2006/3 (Sheldrick, 2008) using SIRAS (single isomorphous replacement with anomalous signal from iodine; Dauter *et al.*, 2000). The $X1_2$ structure was solved with *SHELXC*/*D*/*E* v.2006/3 using single-wavelength anomalous diffraction (SAD; Wang, 1985). The structure of $X1_1$-$X1_2$ was solved with the *MOLREP* molecular-replacement program (v.10.2.23; Vagin & Teplyakov, 2010) using the structures of $X1_1$ and $X1_2$ as models. For all three structures, *ARP*/*wARP* v.7.0 (Cohen *et al.*, 2008) and *Coot* v.0.6-pre-1 (Emsley & Cowtan, 2004) were used for multiple cycles of automatic and manual model building. Further refinement and manual correction was performed using *REFMAC* v.5.5.01 (Murshudov *et al.*, 2011) and *Coot* v.0.6-pre-1 (Emsley & Cowtan, 2004). Eight TLS groups generated using the *TLS Motion Determination* server (Painter & Merritt, 2006*a*,*b*) were used in the final cycle of refinement of the $X1_1$ structure. $X1_2$ was refined with 32 TLS groups and $X1_1$-$X1_2$ with two TLS groups. The resulting structures have been deposited in the Protein Data Bank with PDB codes 3pdg ($X1_1$), 3pe9 ($X1_2$) and 3pdd ($X1_1$-$X1_2$). The data-collection and refinement statistics are shown in Table 1.

Table 2
Pretreatment conditions and compositional analysis of solids used for digestion.

	Hot water-pretreated corn stover	Dilute acid-pretreated corn stover	Alakaline peroxide-pretreated corn stover	Native corn stover	Sigmacel 50
Pretreatment conditions					
Catalyst	None	1% H_2SO_4	2.5% H_2O_2	NA	NA
Temperature (K)	453	433	338	NA	NA
Time (min)	6.5	6.5	60	NA	NA
Compositional analysis (%)					
Solids	46	35	10	NA	NA
Glucan	41.6	60.4	59	34.4	98.3
Xylan	26.4	4.1	28	22.8	2.9
Galactan	0	0	0	1.4	0.82
Arabinan	0.8	0	1.9	4.2	1.35
Mannan	0	0	0	0.6	0
Lignin	17.6	24.6	6.2	11	1.04
Ash	4.6	5.6	1	6.1	ND
Protein	ND	ND	ND	2.3	ND
Uronic acid	ND	ND	ND	3.8	ND
Acetate	2.9	0.8	ND	5.6	ND

2.3. X-ray diffraction and structure determination

The programs *Coot* (Emsley & Cowtan, 2004), *PyMOL* (http://www.pymol.org) and *ICM* (http://www.molsoft.com) were used to compare and analyze structures. Ramachandran plot statistics were calculated using *MolProbity* (Chen *et al.*, 2010) and root-mean-square deviations (r.m.s.d.s) of bond lengths and angles were calculated from ideal values of Engh and Huber stereochemical parameters (Engh & Huber, 1991). The Wilson *B* factor was calculated using *CTRUNCATE* v.1.0.11 (Winn *et al.*, 2011). Structural similarity searches were performed using pairwise secondary-structure matching by *PDBefold* (Krissinel & Henrick, 2004).

2.4. Circular-dichroism (CD) and fluorescence emission spectrum measurements

2.4.1. CD methods. CD measurements were carried out using a Jasco J-715 spectropolarimeter with a jacketed quartz cell with a 1.0 mm path length. The cell temperature was controlled to within ±0.1 K by circulating 90% ethylene glycol through the CD cell jacket using a NESLAB R-111m water bath (NESLAB Instruments, Portsmouth, New Hampshire, USA). The results were expressed as mean residue ellipticity $[\theta]_{MRW}$. The spectra obtained were averages of five scans. The spectra were smoothed using an internal algorithm in the Jasco J-715 software package for Windows. Protein samples were studied in 20 m*M* sodium acetate buffer pH 5.0 with 100 m*M* NaCl at a protein concentration of 0.35 mg ml^{-1} for the near-UV CD. Thermal denaturation of different constructs was monitored by CD in the near-UV (190–260 nm) region. For analysis of thermostability, the temperature was increased from 328 to 378 K with a step size of 0.2 K and monitored at a wavelength of 222 nm.

2.4.2. Fluorescence methods. Steady-state fluorescence measurements were performed on a Jobin Yvon Fluoromax3 spectrofluorometer. Excitation of the protein at 295 nm was selected for tryptophan. Emission spectra were collected from 310 to 400 nm. The $X1_1$-$X1_2$ samples were analyzed at a concentration of 0.1 mg ml^{-1} in 20 m*M* acetate buffer pH 5.0 with 100 m*M* NaCl buffer. Various ligands were added to the following final concentrations: xylotetraose (4 m*M*), xylohexaose (4 m*M*), cellobiose (8 m*M*), cellotetraose (2 m*M*) and cellohexaose (2 m*M*).

2.5. Digestion methods

2.5.1. Effect of temperature on activity. CBM4-Ig-GH9-$X1_1$-$X1_2$ and CBM4-Ig-GH9 constructs were loaded at a concentration of 40 mg protein per gram of glucan acting on a standard cellulose substrate, Sigmacel 50 (Sigma–Aldrich, St Louis, Missouri, USA), with a cellulose loading of 2.4 mg ml^{-1}. Assays were carried out at both 348 and 333 K in 20 m*M* acetate pH 5.0 containing 10 m*M* $CaCl_2$ and 100 m*M* NaCl. To convert cellobiose to glucose at the end of the assay, a small amount of *Aspergillus niger* β-glucosidase (chromatographically purified from the commercial mixture Novozym 188; Novozymes North America, Franklinton, North Carolina, USA) at a concentration of 1.0 mg protein per gram of cellulose substrate was added and allowed to react at 313 K for 1 h. Assays were performed in triplicate and the final glucose concentration was determined using a glucose oxidase assay (Megazyme Ireland).

2.5.2. The effect of X1 modules on substrate digestions. Digestions to test the possible role of X1 modules in cellulose disruption were performed on several different substrates, which included Sigmacel 50 and corn stover pretreated using several different methods (dilute acid, alkaline peroxide and hot water). Corn-stover pretreatment conditions are further described in Table 2. To provide a basis for the maximum theoretical sugar yield achievable from each substrate during enzymatic hydrolysis, portions of each of the pretreated solid samples were dried and subjected to the standard two-stage sulfuric acid hydrolysis method for determining structural carbohydrates in lignocelluloses as described by Sluiter *et al.* (2006). In this method, the carbohydrate content of each pretreated sample was calculated from the carbohydrates released. Enzyme cocktails were then loaded on a milligram enzyme per gram of glucan basis: Spezyme CP (10 mg g^{-1}, Genencor), X1 modules (10 mg g^{-1}), bovine serum albumin (10 mg g^{-1}) or lysozyme (10 mg g^{-1}). Lysozyme was used as a negative control. Digestions were run for 7 d at 323 K and

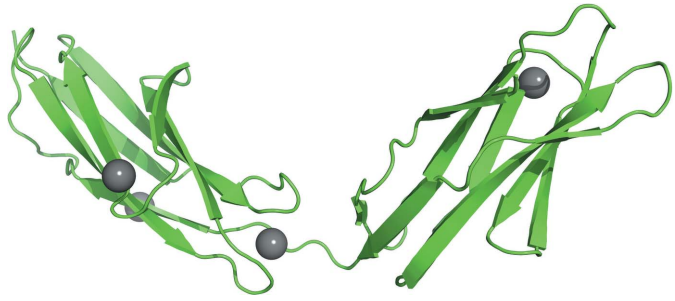

Figure 1
The overall structure of $X1_1$-$X1_2$ with the three Ca atoms shown. β-Strands and loops are shown in green and Ca atoms as gray spheres.

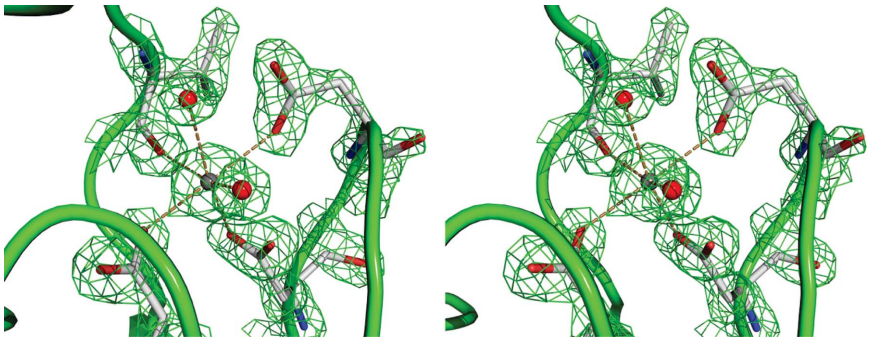

Figure 2
The electron-density map of the Ca atom coordinated between the two X1 modules of the $X1_1$-$X1_2$ structure. This $2F_o - F_c$ map was calculated at 2.0σ after one cycle of *REFMAC*5 (Murshudov *et al.*, 2011). The electron-density map is shown only around the waters and residues coordinated to the Ca atom. Residues are shown in stick representation with O atoms in red, N atoms in blue and C atoms in gray. The calcium is shown as a gray ball and coordinated waters are shown as red balls.

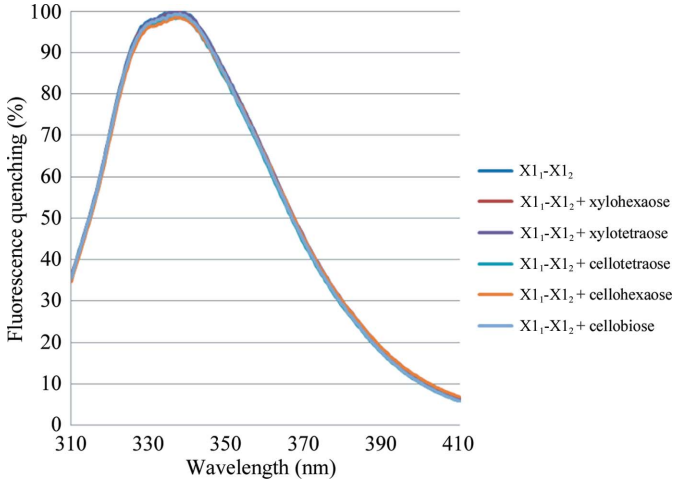

Figure 3
$X1_1$-$X1_2$ substrate interactions monitored by fluorescence spectroscopy. The fluorescence spectra of $X1_1$-$X1_2$ alone and in the presence of cellooligomers and xylooligomers monitored by tryptophan fluorescence spectroscopy are shown.

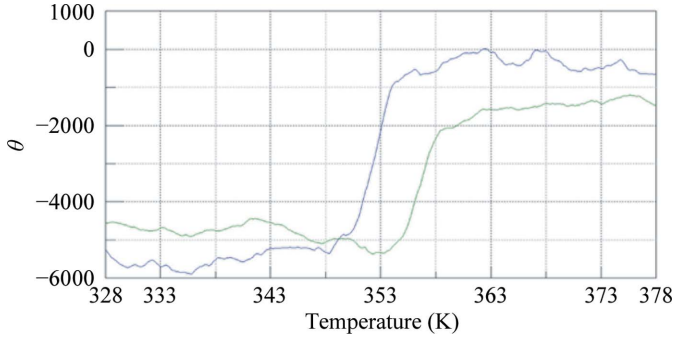

Figure 4
Thermal denaturation of CBM4-Ig-GH9-$X1_1$-$X1_2$ and CBM4-Ig-GH9. Circular-dichroism thermal denaturation curves of CbhA with the X1 modules (green) and without the X1 modules (blue) are shown. The transition midpoints of the curves indicate thermal unfolding of the proteins at 357 and 362 K, respectively.

sugar release was monitored at various time points using a glucose oxidase assay (Megazyme Ireland).

2.6. Simulation methods

All simulations in this section were carried out using the program *PMEMD* from Amber 11 and the parameter set parm99SB (Hornak *et al.*, 2006; Wang *et al.*, 2000). The proteins were solvated in a truncated octahedral box of TIP3P water molecules extending to 15 Å from the surface of the protein to account for the flexibility of the linker. A simulation time step of 2 fs was used together with *SHAKE* (Ryckaert *et al.*, 1977) to constrain covalent bonds between heavy and H atoms. The particle-mesh Ewald method was used together with a nonbonded cutoff of 12 Å. The calcium ions were kept in their original positions from the PDB files and the parameters used for the calcium ions were taken from Aqvist (1990). After equilibration, several 25 ns runs of unconstrained MD were performed for dynamic sampling of states. All analysis was performed with the *ptraj* module of *AmberTools* (Macke *et al.*, 2009; Macke & Case, 1998).

3. Results

3.1. Crystal structures of the CbhA X1 modules

The structure of the *C. thermocellum* CbhA $X1_1$-$X1_2$ module was refined to a resolution of 1.72 Å with R and R_{free} values of 0.167 and 0.226, respectively, and one molecule in the asymmetric unit. The $X1_1$ structure had a resolution of 1.78 Å, with R and R_{free} values of 0.230 and 0.271, respectively, and one molecule in the asymmetric unit. The structure of $X1_2$ had four molecules in the asymmetric unit with a resolution of 1.69 Å and R and R_{free} values of 0.182 and 0.229, respectively.

Both X1 modules of CbhA have the same fold, with two antiparallel β-sheets containing three and four β-strands (Fig. 1). The $X1_1$ structure has two sodium ions, $X1_2$ has 11 I atoms and $X1_1$-$X1_2$ has four Ca atoms modeled (the crystallization conditions contained 0.2 *M* $CaCl_2$). Three of the four Ca atoms in $X1_1$-$X1_2$ are well coordinated, with two of them on the surface and one in the interface between the modules (Fig. 2). The fact that no Ca atoms were observed in the $X1_1$ and $X1_2$ structures indicates that they are not tightly bound, as biologically significant Ca atoms are normally present even when no calcium has been added to the final solution. It is still possible that the one Ca atom in the interface between the two modules is significant, but we cannot be sure without further studies because the identity of the coordinated atoms was only decided on the basis of electron density and coordination. The two sodium ions in the $X1_1$ module were assigned because calcium showed negative density even with reasonably lowered occupancy and contacts ruled out water.

3.2. Structure comparisons

Pairwise secondary-structure matching of structures with at least 70% secondary-structure similarity using *PDBeFold* (Krissinel & Henrick, 2004) found 325 unique structural matches for $X1_1$, 585 matches for $X1_2$ and no matches for $X1_1$-$X1_2$. Similar structures included X1-like (or fibronectin III-like) modules from various sources (including mammalian and fungal, with the ten highest scoring hits) with sequence identities of 21% or less. According to *PDBeFold*, $X1_1$ and $X1_2$ are practically identical, with a root-mean-square deviation of 1.272 Å. However, their sequence similarity is only 25%, which is only slightly better than a number of other bacterial, mammalian and fungal modules that came up in the search. This indicates that they are similar only in protein fold, which rules out any functional comparisons. Also, the $X1_1$ and $X1_2$ modules have just 17 identical residues, excluding the N- and C-terminal ends, suggesting that these modules can tolerate many mutations as long as the protein fold is not affected.

3.3. Biochemical characterization

3.3.1. $X1_1$-$X1_2$ fluorescence measurements. The $X1_1$-$X1_2$ module contains one semi-buried tryptophan residue (Trp148). Tryptophan quenching depends on the proximity and accessibility of the fluorophore to the quenching agent, in this case water; tryptophan is also highly sensitive to its quenching environment and even small conformational changes far away from the actual interaction site can cause large changes in tryptophan fluorescence (Yan & Marriott, 2003; Lakowicz, 2006; Chattopadhyay & Raghuraman, 2004). The lack of an observable fluorescence intensity shift when using both cellooligomers and xylooligomers at millimolar concentrations indicates that there is no direct interaction of the tryptophan, nor a conformational change within the $X1_1$-$X1_2$ module, in the presence of these substrates (Fig. 3). Typically, CBM binding affinities to cellulose are in the micromolar range and large fluorescence intensity shifts are frequently observed when a polysaccharide-binding module, such as CBM4, interacts with a substrate (Alahuhta *et al.*, 2010; Lehtiö *et al.*, 2003). Furthermore, it seems unlikely that an interaction with longer cellulose chains or xylose chains would occur if the $X1_1$-$X1_2$ module cannot bind to short cellooligomers or xylooligomers.

3.3.2. Circular dichroism (CD). We collected CD thermal denaturation spectra, which indicate that CbhA with $X1_1$-$X1_2$ modules (T_m = 357 K) is clearly more thermostable compared with a CbhA construct without these modules (T_m = 352 K) (Fig. 4; Nakanishi *et al.*, 1994). This result clearly shows that the X1 modules play a role in the overall thermostability of CbhA.

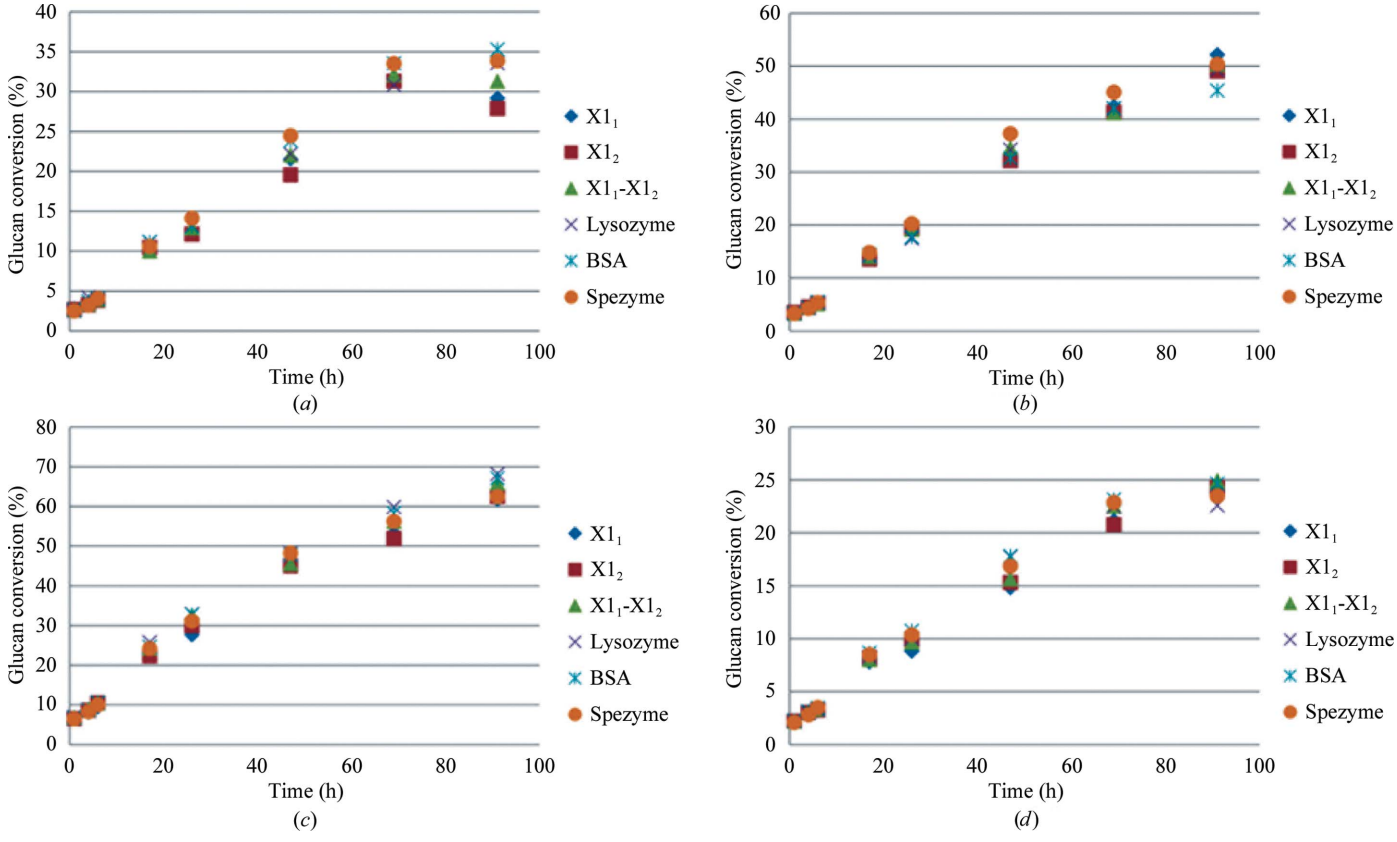

Figure 5
Digestion of model and natural substrates in the presence and absence of X1 modules. Digestion of (*a*) Sigmacel 50, (*b*) dilute acid-pretreated corn stover, (*c*) alkaline peroxide-treated corn stover and (*d*) hot water-pretreated corn stover by Spezyme CP in the presence and absence of X1 modules.

3.4. Digestions using X1 modules

3.4.1. X1 modules do not improve the extent of conversion when added to commercial enzyme cocktails. To test the idea that X1 modules might act as cellulose-disruptor proteins, as suggested by others (Kataeva *et al.*, 2002), we tested a variety of X1 modules in conjunction with the commercial cellulase preparation Spezyme CP to determine whether an improvement in either the rate or the overall extent of conversion of either model compounds or lignocellulosic biomass was possible with the addition of the X1 modules.

We observed that with both the model and process-relevant substrates that we tested, the addition of the purified X1 modules to the Spezyme CP cocktail did not result in a statistically significant difference in either the rate of conversion or the overall conversion of cellulose present compared with either Spezyme CP alone or BSA or lysozyme enzyme controls (Fig. 5). Furthermore, we also observed no effect on xylan conversion under these conditions for any of the substrates (data not shown). These observations held true for both the initial and longer digestions. In some cases, the X1 modules were detrimental to achieving the maximum extent of conversion (Fig. 5*a*).

3.4.2. X1 modules preserve activity at high temperature. To test whether or not the increase in melting temperature observed by CD had a positive effect on the activity of CBM4-Ig-GH9-X1$_1$-X1$_2$, we performed endpoint digestions at two temperatures: 333 K, which mimics the growth temperature, and an elevated condition, 348 K. The endpoint digestion results (Fig. 6) showed that there was an improvement in the extent of conversion compared with the construct without X1 modules of ∼10% at 333 K and ∼40% at 348 K.

3.5. Simulations

The crystal structure of the X1$_1$-X1$_2$ module on its own is not a reliable source of information on the interactions between the two X1 modules or their relative orientation because of crystal contacts. To determine the possible interactions between the two modules, we ran four 25 ns simulations of the duplex. Over the course of these simulations, there were no visually clear interactions between the two modules. Also, as shown in Fig. 7, the correlation of motion between the X1 modules is limited and only exists because of the presence of the linker. On the other hand, the correlation of motion within each module is much more pronounced. This shows that at least in the absence of substrate the X1 modules do not form a compact core but rather a disjointed bi-modular entity. Interestingly, the three well coordinated Ca atoms, including the atom between the two modules, remained bound throughout the simulation.

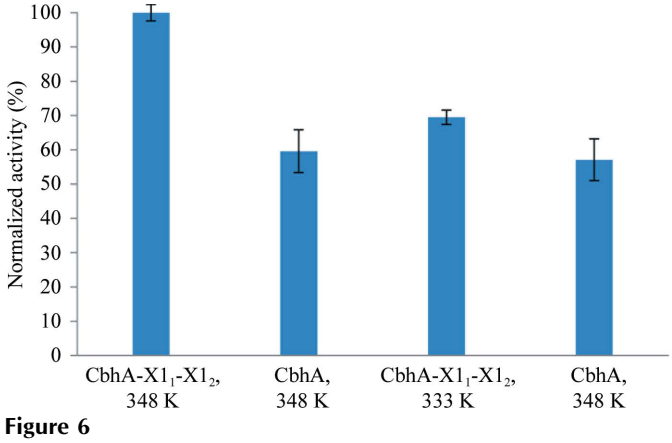

Figure 6
Comparison of the relative extents of conversion of CBM4-Ig-GH9-X1$_1$-X1$_2$ and CBM4-Ig-GH9 at different temperatures (348 and 333 K).

4. Conclusions

In this study, we have described the X-ray diffraction-derived structures of *C. thermocellum* CbhA X1 modules and propose that a possible role of these modules is that of structural stabilizers of the CbhA enzyme complex. Our CD measurements show a drop of 5 K in the thermal denaturation temperature of CbhA without these modules. This clear difference is consistent with the concept proposed by Kataeva *et al.* (2005), who suggested that these modules aid in the refolding of the CbhA catalytic module (GH9) after thermal denaturation. Our activity studies also suggest that one role of the X1 modules is to stabilize the CbhA complex at higher temperatures, which provides a significant (40%) activity enhancement. The work of Adams and coworkers also suggests that a different X module elsewhere in the cellulosome also directly interacts with the cohesin CohI$_9$, forming substantial van der Waals contacts between these modules (Adams *et al.*, 2010). Additionally, whereas we have no direct evidence for interaction of the X1 modules and the entire CbhA complex, such an interaction may indeed occur.

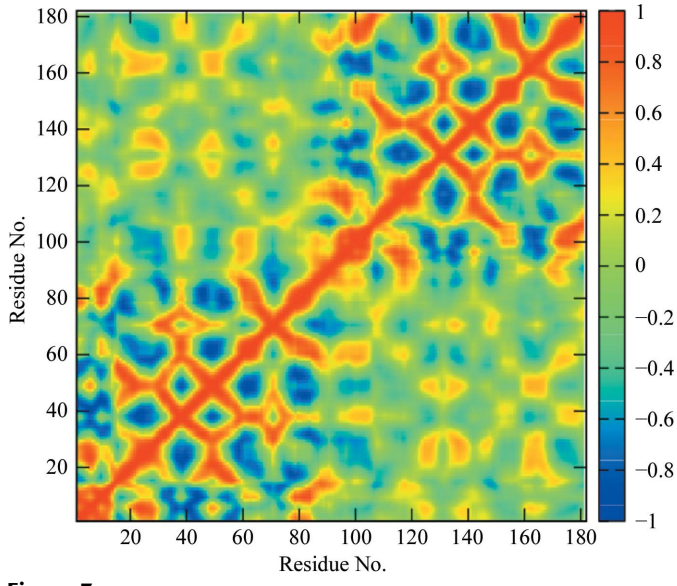

Figure 7
Residue cross-correlation map of two X1 modules. A value of 1 indicates correlation of motion, whereas −1 is indicative of anticorrelation of motion and zero represents a total lack of motion correlation. This map was calculated using the average of four independent 25 ns molecular-dynamics simulations.

It has also been suggested previously (Kataeva *et al.*, 2002) that X1 modules might act as disruptor-type proteins (Harris *et al.*, 2010). We have explored this possibility and have presented two results which suggest that this idea seems to be unlikely. Firstly, we show that the CbhA X1 modules alone do not interact with cellodextrins or xylodextrins in a biologically relevant manner. Furthermore, we have also shown that the free X1 modules do not improve the digestibility of pretreated corn-stover biomass. Such synergism has recently been reported for oxidative enzymes from the GH61 family (Harris *et al.*, 2010). Therefore, whereas the X1 modules do not appear to interact with or modify the surface of biomass as has been previously suggested, we have shown that they are important for the thermostability of the CbhA multimodule enzyme, preserving its activity at higher temperatures.

This work was supported by the DOE Office of Science, Office of Biological and Environmental Research through the BioEnergy Science Center (BESC), a DOE Bioenergy Research Center. Computational time for this research was supported in part by the Golden Energy Computing Organization at the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory. Simulations were also performed in part using the Texas Advanced Computing Center Ranger cluster under the National Science Foundation Teragrid grant No. TG-MCB090159.

References

Adams, J. J., Currie, M. A., Ali, S., Bayer, E. A., Jia, Z. & Smith, S. P. (2010). *J. Mol. Biol.* **396**, 833–839.
Alahuhta, M., Xu, Q., Bomble, Y. J., Brunecky, R., Adney, W. S., Ding, S.-Y., Himmel, M. E. & Lunin, V. V. (2010). *J. Mol. Biol.* **402**, 374–387.
Aqvist, J. (1990). *J. Phys. Chem.* **94**, 8021–8024.
Bayer, E. A., Belaich, J. P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.
Chattopadhyay, A. & Raghuraman, H. (2004). *Curr. Sci.* **87**, 175–180.
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.
Chiriac, A. I., Cadena, E. M., Vidal, T., Torres, A. L., Diaz, P. & Pastor, F. I. (2010). *Appl. Microbiol. Biotechnol.* **86**, 1125–1134.
Cohen, S. X., Ben Jelloul, M., Long, F., Vagin, A., Knipscheer, P., Lebbink, J., Sixma, T. K., Lamzin, V. S., Murshudov, G. N. & Perrakis, A. (2008). *Acta Cryst.* D**64**, 49–60.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
Devillard, E., Goodheart, D. B., Karnati, S. K., Bayer, E. A., Lamed, R., Miron, J., Nelson, K. E. & Morrison, M. (2004). *J. Bacteriol.* **186**, 136–145.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Gold, N. D. & Martin, V. J. (2007). *J. Bacteriol.* **189**, 6787–6795.
Harris, P. V., Welner, D., McFarland, K. C., Re, E., Navarro Poulsen, J. C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S. & Lo Leggio, L. (2010). *Biochemistry*, **49**, 3305–3316.
Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. (2006). *Proteins*, **65**, 712–725.
Jee, J. G., Ikegami, T., Hashimoto, M., Kawabata, T., Ikeguchi, M., Watanabe, T. & Shirakawa, M. (2002). *J. Biol. Chem.* **277**, 1388–1397.
Jindou, S., Petkun, S., Shimon, L., Bayer, E. A., Lamed, R. & Frolow, F. (2007). *Acta Cryst.* F**63**, 1044–1047.
Kataeva, I. A., Brewer, J. M., Uversky, V. N. & Ljungdahl, L. G. (2005). *FEBS Lett.* **579**, 4367–4373.
Kataeva, I. A., Seidel, R. D. III, Shah, A., West, L. T., Li, X.-L. & Ljungdahl, L. G. (2002). *Appl. Environ. Microbiol.* **68**, 4292–4300.
Kataeva, I. A., Uversky, V. N., Brewer, J. M., Schubot, F., Rose, J. P., Wang, B.-C. & Ljungdahl, L. G. (2004). *Protein Eng. Des. Sel.* **17**, 759–769.
Kataeva, I. A., Uversky, V. N. & Ljungdahl, L. G. (2003). *Biochem. J.* **372**, 151–161.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Lakowicz, J. R. (2006). *Principles of Fluorescence Spectroscopy*. New York: Springer.
Lehtiö, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M. & Teeri, T. T. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 484–489.
Little, E., Bork, P. & Doolittle, R. F. (1994). *J. Mol. Evol.* **39**, 631–643.
Macke, T. A. & Case, D. A. (1998). *Molecular Modeling of Nucleic Acids.*, 1st ed., edited by N. B. Leontes & J. SantaLucia Jr, pp. 379–393. Washington DC: American Chemical Society.
Macke, T., Svrcek Seiler, W. A., Brown, R. A., Kolossvary, I., Bomble, Y. J. & Case, D. A. (2009). *The NAB Molecular Manipulation Language*. http://casegroup.rutgers.edu/casegr-sh-2.2.html.
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.* D**67**, 355–367.
Nakanishi, K., Berova, N. & Woody, R. (1994). *Circular Dichroism: Principles and Applications.* New York: VCH.
Painter, J. & Merritt, E. A. (2006*a*). *Acta Cryst.* D**62**, 439–450.
Painter, J. & Merritt, E. A. (2006*b*). *J. Appl. Cryst.* **39**, 109–111.
Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). *J. Comput. Phys.* **23**, 327–341.
Schubot, F. D., Kataeva, I. A., Chang, J., Shah, A. K., Ljungdahl, L. G., Rose, J. P. & Wang, B.-C. (2004). *Biochemistry*, **43**, 1163–1170.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sluiter, A., Hames, B., Ruiz, R., Scarlata, C., Sluiter, J., Templeton, D. & Crocker, D. (2006). *Determination of Structural Carbohydrates and Lignin in Biomass*. Technical Report NREL/TP-510-42618. Golden: National Renewable Energy Laboratory.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Wang, J. M., Cieplak, P. & Kollman, P. A. (2000). *J. Comput. Chem.* **21**, 1049–1074.
Watanabe, T., Ito, Y., Yamada, T., Hashimoto, M., Sekine, S. & Tanaka, H. (1994). *J. Bacteriol.* **176**, 4465–4472.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yan, Y. & Marriott, G. (2003). *Curr. Opin. Chem. Biol.* **7**, 635–640.
Zverlov, V. V., Velikodvorskaya, G. V., Schwarz, W. H., Bronnenmeier, K., Kellermann, J. & Staudenbauer, W. L. (1998). *J. Bacteriol.* **180**, 3091–3099.