

The novel thermostable cellulose-degrading enzyme DtCel5H from *Dictyoglomus thermophilum*: crystallization and X-ray crystallographic analysis

Flavia Squeglia, Rita Berisio and Alessia Ruggiero*

Institute of Biostructures and Bioimaging (IBB-CNR), Italy. *Correspondence e-mail: alessia.ruggiero@unina.it

Received 22 September 2017

Accepted 21 November 2017

Edited by R. L. Stanfield, The Scripps Research Institute, USA

Keywords: crystal structure; endoglucanase; *Dictyoglomus thermophilum*; thermostability; biomass; renewable energy.

Supporting information: this article has supporting information at journals.iucr.org/f

Cellulose-based products constitute the great majority of municipal waste, and applications of cellulases in the conversion of waste biomass to biofuels will be a key technology in future biorefineries. Currently, multi-enzymatic pre-treatment of biomass is a crucial step in making carbohydrates more accessible for subsequent fermentation. Using bioinformatics analysis, endo- β -(1,4)-glucanase from *Dictyoglomus thermophilum* (DtCel5H) was identified as a new member of glycosyl hydrolase family 5. The gene encoding DtCel5H was cloned and the recombinant protein was overexpressed for crystallization and biophysical studies. Here, it is shown that this enzyme is active on cellulose substrates and is highly thermostable. Crystals suitable for crystallographic investigations were also obtained in different crystallization conditions. In particular, ordered crystals of DtCel5H were obtained using either ammonium sulfate or polyethylene glycol (PEG) as a precipitant agent. The crystals obtained in the presence of ammonium sulfate belonged to space group $P3_2$, with unit-cell parameters $a = 73.1$, $b = 73.1$, $c = 127.8$ Å, and diffracted to 1.5 Å resolution, whereas the second crystal form belonged to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 49.3$, $b = 67.9$, $c = 103.7$ Å, and diffracted to 1.6 Å resolution. The crystal structure was solved in both space groups using molecular-replacement methods. Structure–activity and structure–stability studies of DtCel5H will provide insights for the design of high-performance enzymes.

1. Introduction

The development of renewable energy is gaining scientific and public attention, leading to a shift from depleting and un-sustainable fossil fuels to cleaner energy. Among the available renewable energy sources, biomass provides almost 10% of the global bioenergy supply (Scarlat *et al.*, 2015; Ragauskas *et al.*, 2006). Biomass supply is categorized into three sectors: agriculture, forestry and wastes (Searle & Malins, 2015). All of these sources of biofuel are largely made up of cellulose and hemicellulose (70–80%; Jørgensen *et al.*, 2007). Cellulose is a homo-polysaccharide composed of β -D-pyranose units linked by β -(1,4)-glycosidic bonds, and is the main component of plant cell walls. Hemicellulose is present as the matrix that surrounds the cellulose skeleton, and is made up of several heteropolymers, such as arabinoxylans. The composition of cell walls in plants varies between species and may depend on the cell type and developmental stage (Somerville *et al.*, 2004). Other protective polymers such as lignin, suberin or cutin are anchored to or embedded in plant cell walls. Furthermore, plants such as algae possess cell walls made of glycoproteins and polysaccharides such as carrageenan and agar that are absent from land plants (Ruggiero *et al.*, 2016).



© 2018 International Union of Crystallography

Currently, pretreatment of biomass is an important tool for cellulose-based conversion processes (Kumar *et al.*, 2009). This step is essential to change the structure of cellulosic biomass and to make the cellulose more available to the enzymes that convert the carbohydrate polymers into fermentable sugars (Mosier *et al.*, 2005). Several studies have reported pretreatment methods, such as biological, chemical, mechanical and thermal processes, as well as their combinations, to enhance biomass decomposition (Wagner *et al.*, 2013; Us & Perendeci, 2012; Horn *et al.*, 2012). The use of cellulases for converting cellulose into oligomers and sugar monomers is denoted enzymatic saccharification; an effective biological pretreatment employs a combination of enzymes that can work synergistically to expand small pores and increase access by opening the cell-wall matrix (Jeremic *et al.*, 2014; Amin *et al.*, 2017). Indeed, owing to the complex chemical structure of the plant cell wall, distinct enzymes are often needed for the degradation of lignocellulosic materials into sugars (Liao *et al.*, 2011). For instance, cellulose is hydrolyzed by a mixture of cellulases (endoglucanases and exoglucanases, β -glucosidases) and hemicellulose is hydrolyzed by the action of different hemicellulases (for example xylanases and mannanases; Brummer *et al.*, 2014).

Both bacteria and fungi are able to produce extracellular cellulases that are either free or cell-associated to hydrolyze and metabolize insoluble cellulose (Kuhad *et al.*, 2011). Most cellulase enzymes are relatively unstable at high temperatures. Therefore, cellulases from extremophiles with high stability are expected to be useful for the industrial hydrolysis of plant biomass during its processing into biofuels (Turner *et al.*, 2007; Maki *et al.*, 2009).

Dictyoglomus thermophilum is an extremely thermophilic cellulolytic anaerobe that was originally isolated from a hot spring in Japan (Saiki *et al.*, 1985). It is generally adapted to living at high temperatures, with an optimum growth temperature of between 73 and 78°C and a maximum temperature of 80°C. This organism is of interest since it produces xylanases (Li *et al.*, 2013; Zhang *et al.*, 2010) and α -amylases (Fukusumi *et al.*, 1988; Horinouchi *et al.*, 1988). The availability of the entire genome sequence of *D. thermophilum* (Coil *et al.*, 2014) has provided the chance to discover new glycosidases with cellulase or hemicellulase activities (Shi *et al.*, 2013).

In this context, here we report the recombinant production, crystallization and X-ray crystallographic analysis of a novel enzyme endowed with a strong cellulose activity at high temperature. These studies have great potential for biotechnological applications.

2. Materials and methods

2.1. Macromolecule production

The gene encoding DtCel5H, deprived of its first 29 amino acids (which are predicted to form a transmembrane helix) and with its two cysteine codons mutated, was synthesized by GENEWIZ (Sigma–Aldrich, UK) and subsequently

Table 1
Macromolecule-production information.

Source organism	<i>D. thermophilum</i>
Forward primer 1	AAACTGGAATGCTCCCGATTCTTCGAGGC ATTAACATGG
Forward primer 2†	CATGCCATGGGCTCTTCTATGAATTTAAA CTGGAATGCTCCCGAT
Reverse primer‡	GATGCTCGAGCTTAGTTTCTGGTATTAGAG C
Expression vector	pETM-13
Expression host	<i>E. coli</i> BL21(DE3) or <i>E. coli</i> Rosetta2(DE3)
Complete amino-acid sequence of the construct produced	MGSSYEFKLEMLPILRGINMGNALEAPREG EWGVVVKDEYFKIIEKAGFQSHVRIPIRW NAHADMNPPYTIKAFKFFDRVDHVVNEAL KNELFVIINIHYYEIMQYPEKHKDRFL ALWRQIAEHYRDYPTLLIFELLNEPSMN LTANLWNKYLAETIKVIRESNPDRYIVV GPNWNSIYALKDLKLPKDEKNIIVTFH YYNPFSTHQAEWVQSPFPVGVRRWGR EDEKAEIDRELDMDLWDSIRNGKVPPLYM GEFGAYSKADMSRVRWTSYVARSAEKR GIAWAYWEPFAGFGAFDPVKNEWRKPLL KALIPETKLEHHHHHH

subcloned into the expression vector pETM-13 (EMBL, Germany) between the NcoI and XhoI restriction sites. DtCel5H was expressed with a C-terminal histidine tag in different strains of *Escherichia coli* under varied experimental conditions (*e.g.* temperature, IPTG concentration and culture medium). Subsequently, a longer construct containing an extra N-terminal portion of 13 amino acids was added to the protein (sequence MGSSYEFKLEMLP) by two consecutive overlapping steps of PCR. For these reactions the primers reported in Table 1 were adopted. The gene was inserted into the pETM-13 vector between the NcoI and XhoI sites. This protein construct was successfully overexpressed in *E. coli* BL21 (DE3) cells. Briefly, an overnight starting culture of 10 ml was prepared for growth in 1 l Luria–Bertani (LB) medium containing 50 $\mu\text{g ml}^{-1}$ kanamycin, which was then induced with 1 mM IPTG at 22°C for 16 h. Expression details are given in Table 1.

DtCel5H was purified by sonicating cells resuspended in binding buffer (50 mM Tris–HCl, 10 mM imidazole, 300 mM NaCl pH 8.0) containing a protease-inhibitor cocktail (Roche Diagnostics). The lysate was cleared by centrifugation at 18 000g and the supernatant was loaded onto 5 ml Ni–NTA resin (Qiagen) equilibrated with binding buffer. After washing with ten volumes of binding buffer, the protein was eluted by adding 200 mM imidazole to the binding buffer. The fractions containing the eluted protein were pooled, concentrated and then loaded onto a Superdex 200 HR 16/60 gel-filtration column (GE Healthcare) equilibrated with 50 mM Tris–HCl, 200 mM NaCl pH 8.0 for a further purification step. The protein eluted in a single peak and was homogeneous, as judged by SDS–PAGE analysis. The protein was concentrated using a centrifugal filter (Merck Millipore) and the concentration was determined by UV absorbance using an ϵ value of 95 340 $\text{M}^{-1} \text{cm}^{-1}$.

The molar mass of DtCel5H was determined by multi-angle light scattering (MALS). The purified protein was reloading

Table 2
Data collection and processing.

Values in parentheses are for the highest resolution shell.

	Crystal form I	Crystal form II
Diffraction source	Cu $K\alpha$	Cu $K\alpha$
Wavelength (Å)	1.542	1.542
Temperature (K)	100	100
Detector	Saturn 944 CCD	Saturn 944 CCD
Crystal-to-detector distance (mm)	50	50
Rotation range per image (°)	0.5	0.5
Total rotation range (°)	180	180
Exposure time per image (s)	10	10
Space group	$P3_2$	$P2_12_12_1$
a, b, c (Å)	73.1, 73.1, 127.8	49.3, 67.9, 103.7
Mosaicity (°)	0.59	0.3
Resolution range (Å)	50.0–1.5 (1.50–1.48)	50.0–1.6 (1.58–1.55)
Total No. of reflections	306316	171251
No. of unique reflections	124816	50537
Completeness (%)	92.1 (75)	89.1 (76)
$\langle I/\sigma(I) \rangle$	34.9 (3.2)	29.5 (2.1)
$R_{\text{r.i.m.}}$ (%)	5.5 (31.6)	6.4 (44.7)

onto a Superdex 200 HR 10/30 gel-filtration column (GE Healthcare) equilibrated with 20 mM Tris–HCl, 200 mM NaCl pH 8.0. An ÄKTA chromatography system (GE Healthcare) was used coupled to a MALS detector (mini-DAWN TREOS) and a refractive-index detector (Optilab DSP; Wyatt Technology, Woburn, Massachusetts, USA).

2.1.1. Native PAGE zymography. For native PAGE, different amounts of purified DtCel5H (20–60 µg per lane) were mixed with a buffer consisting of 25 mM Tris–HCl, 10% (w/v) glycerol pH 8.8. The samples, without prior heating, were loaded onto a 10% gel containing 0.1% (w/v) carboxymethyl cellulose (CMC; low viscosity; catalogue No. C5678, Sigma–Aldrich, Italy). Electrophoresis was performed at room temperature in a vertical mini-gel system (Mini-PROTEAN II, Bio-Rad, USA) at a constant current of 30 mA for 2.5 h. The native gel was carefully removed from the electrophoretic chamber and cut into two slices for endoglucanase-activity staining. One gel slice was stained using the Coomassie staining method and the other was soaked in 50 mM sodium citrate–phosphate, 300 mM NaCl pH 4.7 at 50°C for 2 h prior to staining with 0.2% (w/v) Congo red. This portion of the gel was destained with 2 M NaCl until a clear zone was observed against the red background.

2.2. Crystallization

Freshly concentrated protein (at 37 mg ml⁻¹) was used for crystallization experiments. Initial crystallization trials of DtCel5H were carried out by the sitting-drop vapour-diffusion method using a crystallization workstation (Hamilton Robotics). 288 high-throughput reagents (Crystal Screen, Crystal Screen 2, Index, PEG/Ion and PEG/Ion 2; Hampton Research) were tested. All drops were made by mixing 0.2 µl protein solution with an equal volume of reservoir solution on 96-well plates and were equilibrated against 100 µl reservoir solution. Optimization of the crystallization conditions was performed manually by fine-tuning the protein and precipitant concentrations.

2.3. Data collection and processing

High-resolution diffraction data for the ligand-free form of DtCel5H were collected in-house at 100 K using a Rigaku MicroMax-007 HF generator producing Cu $K\alpha$ radiation and equipped with a Saturn 944 CCD detector. In all cases, cryo-protection of the crystals was achieved by rapid soaking (1–2 s) in a solution containing glycerol in the range 15–20% (v/v). An oscillation range of 0.5° and an X-ray dose corresponding to an exposure of about 10 s were adopted for all experiments.

The data sets were scaled and merged using the *HKL-2000* package (Otwinowski & Minor, 1997). Data-collection statistics are listed in Table 2.

In attempt to determine the initial phases, molecular replacement was performed with *Phaser* (McCoy *et al.*, 2007) using the coordinates of the endoglucanase Cel5A from *Thermatoga maritima* as a search model (59% sequence identity for residues 32–330; PDB entry 3mmu; Pereira *et al.*, 2010).

3. Results and discussion

By comparative sequence analyses with *BLASTP* using the endoglucanase Cel5A (EC 3.2.1.4) from *T. maritima* (UniProtKB accession No. Q9X273) as a query, we have

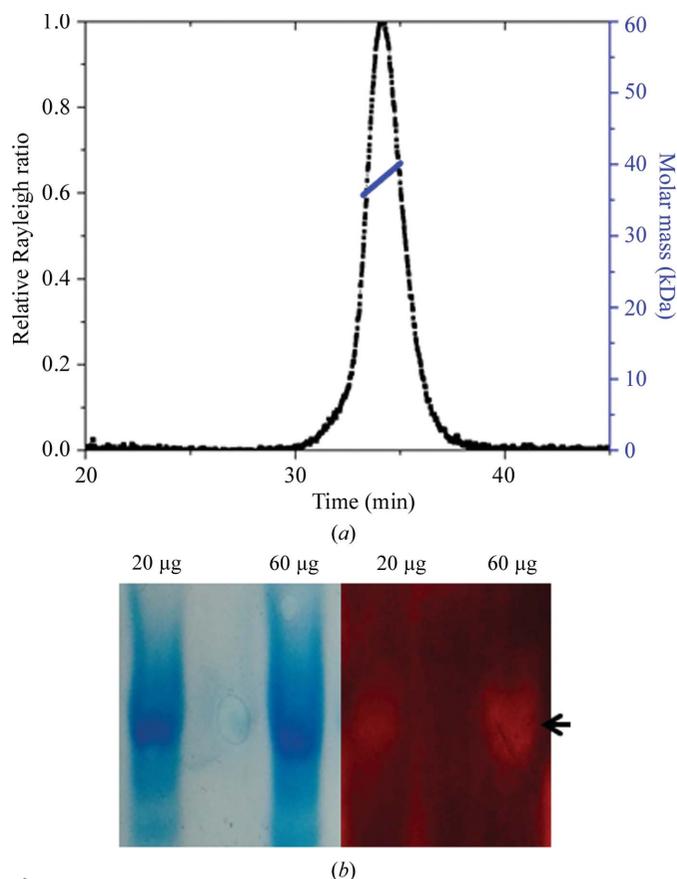


Figure 1
(a) Multi-angle light-scattering result for DtCel5H. The blue line indicates the experimental molecular weight. (b) Native PAGE zymography of DtCel5H. The left panel shows the Coomassie-stained gel, whereas the right panel shows the in-gel activity after Congo red staining. The black arrow indicates the position of the CMC degradation zone owing to DtCel5H activity.

identified a new uncharacterized protein from *D. thermophilum* (UniProtKB accession No. B5YAS2), here denoted DtCel5H. This protein, encoded by the gene DICTH_0008, contains 335 amino acids and contains a putative N-terminal transmembrane helix (residues 5–21). Therefore, a truncated version of the enzyme starting from the Ile30 residue was rationally engineered for its recombinant expression. The two cysteine residues (Cys156 and Cys308) were also mutated to serines. The gene encoding residues 30–334 of DtCel5H was synthesized and successfully subcloned in pETM-13 vector. Protein production from this construct was difficult to achieve, as most of the protein was recovered in *E. coli* inclusion bodies. To enhance the solubility of DtCel5H, a longer version of the enzyme, starting from Ser18, was prepared using two consecutive overlapping steps of PCR and was cloned in the pETM-13 vector. This protein construct was successfully

expressed in the soluble portion of the *E. coli* lysate. The protein was purified by coupling two consecutive chromatographic steps. An initial step was carried out on an affinity column, and a gel-filtration step was subsequently used to achieve a purity of >99%.

The oligomerization state of DtCel5H was assessed using MALS. The calculated molecular weight of DtCel5H ($37\,890 \pm 114$ Da), including its C-terminal histidine tag, indicates that the protein is monomeric in solution (Fig. 1*a*). Using zymography, we showed that DtCel5H is thermostable and fully active against cellulose-based substrates at 50°C (Fig. 1*b*).

Initial crystallization screening using commercially available solutions revealed promising conditions in the presence of ammonium sulfate. In particular, after 1 d bipyramidal-shaped crystals of DtCel5H were obtained by mixing protein solution at 20 mg ml^{-1} with 0.2 M ammonium sulfate, 5% (w/v)

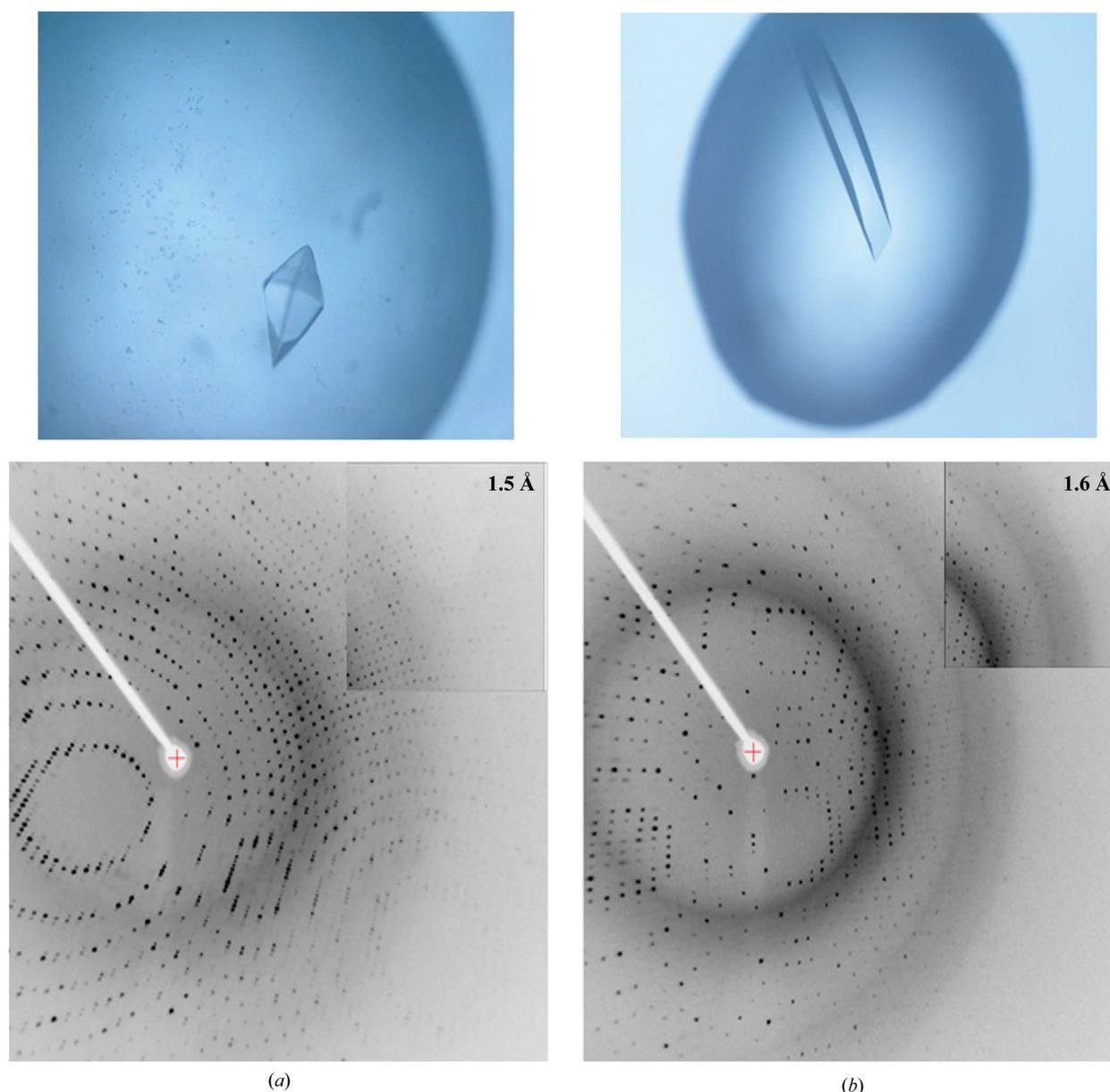


Figure 2 Photographs and diffraction patterns of crystals of DtCel5H obtained using (a) 2.0 M ammonium sulfate (crystal form I) and (b) 18% (w/v) PEG 8K (crystal form II) as precipitants. Expanded views are shown in the insets in (b) to show the maximum resolution of the spots.

2-propanol (Fig. 2a). These crystals diffracted to 1.5 Å resolution and belonged to space group $P3_2$, with unit-cell parameters $a = 73.1$, $b = 73.1$, $c = 127.8$ Å (crystal form I in Table 2). Matthews coefficient calculations suggested the presence of two molecules in the asymmetric unit ($V_M = 2.6$ Å³ Da⁻¹, 52.6% solvent content; Matthews, 1968). Consistent with the V_M calculation, the self-rotation function computed for these data showed the presence of a significant peak (RF/sig 4.34) in the $\chi = 180^\circ$ section (Supplementary Fig. S1).

Preliminary analysis of the electron-density map clearly identified the catalytic pocket of DtCel5H and revealed that this pocket is occluded, in the adopted experimental conditions, by the N-terminal region of an adjacent molecule (data not shown). To allow the binding of glucosidic substrates to this new cellulase, we searched for alternative crystal-packing forms. Crystallization experiments carried out using a more concentrated protein sample (35 mg ml⁻¹) allowed us to find a new crystal form which grew in the presence of 18% (w/v) PEG 8K, 0.2 M calcium acetate, 0.1 M sodium cacodylate pH 6.5 (Fig. 2b). These crystals diffracted to 1.6 Å resolution and belonged to space group $P2_12_12_1$, with unit-cell parameters $a = 49.3$, $b = 67.9$, $c = 103.7$ Å (crystal form II in Table 2). Matthews coefficient calculations suggested the presence of one molecule in the asymmetric unit ($V_M = 2.3$ Å³ Da⁻¹, 46% solvent content). Model building and structure refinement are currently in progress.

This work will allow structural investigations of a fully active new cellulase with thermostable properties for use in biotechnological applications. The structural characterization of this enzyme will provide molecular determinants useful for the design of enzymes with higher cellulase activity.

Acknowledgements

The authors acknowledge Giosuè Sorrentino and Maurizio Amendola for their technical assistance

Funding information

This work was financially supported by MIUR (project PON03PE_00157), SMART GENERATION (Protocollo 0001355) and by FEP Regione Campania 2007–2013 Misura 3.5 (project ‘Use of marine plants to produce removable energy’).

References

Amin, F. R., Khalid, H., Zhang, H., Rahman, S. U., Zhang, R., Liu, G. & Chen, C. (2017). *AMB Express*, **7**, 72.

- Brummer, V., Jurena, T., Hlavacek, V., Omelkova, J., Bebar, L., Gabriel, P. & Stehlik, P. (2014). *Bioresour. Technol.* **152**, 543–547.
- Coil, D. A., Badger, J. H., Forberger, H. C., Riggs, F., Madupu, R., Fedorova, N., Ward, N., Robb, F. T. & Eisen, J. A. (2014). *Genome Announc.* **2**, e00109-14.
- Fukusumi, S., Kamizono, A., Horinouchi, S. & Beppu, T. (1988). *Eur. J. Biochem.* **174**, 15–21.
- Horinouchi, S., Fukusumi, S., Ohshima, T. & Beppu, T. (1988). *Eur. J. Biochem.* **176**, 243–253.
- Horn, S. J., Vaaje-Kolstad, G., Westereng, B. & Eijsink, V. G. (2012). *Biotechnol. Biofuels*, **5**, 45.
- Jeremic, D., Goacher, R. E., Yan, R., Karunakaran, C. & Master, E. R. (2014). *Biotechnol. Biofuels*, **7**, 496.
- Jørgensen, H., Kristensen, J. B. & Felby, C. (2007). *Biofuels Bioprod. Bioref.* **1**, 119–134.
- Kuhad, R. C., Gupta, R. & Singh, A. (2011). *Enzym. Res.* **2011**, 280696.
- Kumar, P., Barrett, D. M., Delwiche, M. J. & Stroeve, P. (2009). *Ind. Eng. Chem. Res.* **48**, 3713–3729.
- Li, H., Kankaanpää, A., Xiong, H., Hummel, M., Sixta, H., Ojamo, H. & Turunen, O. (2013). *Enzyme Microb. Technol.* **53**, 414–419.
- Liao, H., Zhang, X.-Z., Rollin, J. A. & Zhang, Y.-H. P. (2011). *Biotechnol. J.* **6**, 1409–1418.
- Maki, M., Leung, K. T. & Qin, W. (2009). *Int. J. Biol. Sci.* **5**, 500–516.
- 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.
- Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapfel, M. & Ladisch, M. (2005). *Bioresour. Technol.* **96**, 673–686.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Pereira, J. H., Chen, Z., McAndrew, R. P., Sapra, R., Chhabra, S. R., Sale, K. L., Simmons, B. A. & Adams, P. D. (2010). *J. Struct. Biol.* **172**, 372–379.
- Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J. Jr, Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., Murphy, R., Templer, R. & Tschaplinski, T. (2006). *Science*, **311**, 484–489.
- Ruggiero, A., Squeglia, F., Romano, M. & Berisio, R. (2016). *Curr. Biotechnol.* **5**, 296–304.
- Saiki, T., Kobayashi, Y., Kawagoe, K. & Beppu, T. (1985). *Int. J. Syst. Bacteriol.* **35**, 253–259.
- Scarlat, N., Dallemand, J.-F., Monforti-Ferrario, F. & Nita, V. (2015). *Environ. Dev.* **15**, 3–34.
- Searle, S. & Malins, C. (2015). *Glob. Change Biol. Bioenergy*, **7**, 328–336.
- Shi, R., Li, Z., Ye, Q., Xu, J. & Liu, Y. (2013). *Bioresour. Technol.* **142**, 338–344.
- Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredes, A., Persson, S., Raab, T., Vorwerk, S. & Youngs, H. (2004). *Science*, **306**, 2206–2211.
- Turner, P., Mamo, G. & Karlsson, E. N. (2007). *Microb. Cell Fact.* **6**, 9.
- Us, E. & Perendeci, N. A. (2012). *Chem. Eng. J.* **181–182**, 120–131.
- Wagner, A. O., Schwarzenauer, T. & Illmer, P. (2013). *J. Environ. Manage.* **129**, 357–360.
- Zhang, W., Lou, K. & Li, G. (2010). *Appl. Biochem. Biotechnol.* **160**, 1484–1495.