Crystal structure of a putative exo- β -1,3-galactanase from *Bifidobacterium bifidum* S17

Andre S. Godoy,^a Mariana Z. T. de Lima,^a Cesar M. Camilo^b and Igor Polikarpov^{a*}

Received 11 December 2015

Accepted 1 March 2016

Edited by W. N. Hunter, University of Dundee,
Scotland**Keywords:** galactanase; *Bifidobacterium*;
GH43; family 43; hydrolase.**PDB reference:** exo- β -1,3-galactanase, 5flw**Supporting information:** this article has
supporting information at journals.iucr.org/f^aDepartamento de Física em São Carlos, Universidade de São Paulo, Avenida Trabalhador Saocarlene 400, 13560-970 São Carlos-SP, Brazil, and ^bCentro de Tecnologia Canavieira, Fazenda Santo Antonio, PO Box 162, 13400-970 Piracicaba-SP, Brazil. *Correspondence e-mail: ipolikarpov@ifsc.usp.br

Given the current interest in second-generation biofuels, carbohydrate-active enzymes have become the most important tool to overcome the structural recalcitrance of the plant cell wall. While some glycoside hydrolase families have been exhaustively described, others remain poorly characterized, especially with regard to structural information. The family 43 glycoside hydrolases are a diverse group of inverting enzymes; the available structure information on these enzymes is mainly from xylosidases and arabinofuranosidase. Currently, only one structure of an exo- β -1,3-galactanase is available. Here, the production, crystallization and structure determination of a putative exo- β -1,3-galactanase from *Bifidobacterium bifidum* S17 (*BbGal43A*) are described. *BbGal43A* was successfully produced and showed activity towards synthetic galactosides. *BbGal43A* was subsequently crystallized and data were collected to 1.4 Å resolution. The structure shows a single-domain molecule, differing from known homologues, and crystal contact analysis predicts the formation of a dimer in solution. Further biochemical studies are necessary to elucidate the differences between *BbGal43A* and its characterized homologues.

1. Introduction

Growing insecurity about the availability and environmental impact of nonrenewable energy resources has promoted an intensive search for renewable alternatives (Mohanram *et al.*, 2013). In such a scenario, the production of enzymes for the transformation of plant biomass into biofuels and green chemicals has become an important area of research and technological development in many industries and scientific institutions. Carbohydrate-active enzymes (CAZymes; Lombard *et al.*, 2014) are the prime biological tool to overcome the structural recalcitrance of the plant cell wall and reduce its biochemical complexity to simple monosaccharides and oligosaccharides (Payne *et al.*, 2015).

Glycoside hydrolase family 43 (GH43) is a complex group of inverting CAZymes with distinct functions, including β -xylosidases (EC 3.2.1.37), α -L-arabinofuranosidases (EC 3.2.1.55), arabinanases (EC 3.2.1.99), xylanases (EC 3.2.1.8) and galactan 1,3- β -galactosidases (EC 3.2.1.145), among others (Lombard *et al.*, 2014). The family contains 31 structurally characterized members, which are mainly xylosidases and arabinofuranosidases. The exo- β -1,3-galactanase from *Clostridium thermocellum* (*Ct1,3Gal43A*; PDB entry 3vsf) presently remains the only structurally and biochemically characterized GH43 β -1,3-galactanase (Jiang *et al.*, 2012).

Here, we describe the cloning, purification and crystal structure solution of a member of GH43 from *Bifidobacterium*

Table 1

Cloning information for *BbGal43A*.

Source organism	<i>B. bifidum</i> S17
DNA source	Genomic
Forward primer	CAGGGCGCCATGGCTGATAGATTCCGGAGC
Reverse primer	GACCCGACGCGGTTAGCGGTGGGTGGTG
Expression vector	pET_TrxA
Expression host	<i>E. coli</i> Rosetta(DE3)
Complete amino-acid sequence of the construct produced	MADRFGAFLPHDTSGDVAQLHGIGLQKFGDTWYA-YGENKVNGLFQGVCCYTTTDFIAWRSHGIVL-DVQEDGSALAADRIGERPKVLHCPATGKYVMY-IHAETPDYGYAHIGVAVADAPTGPFAFQTITIT-WRGLSRDIGVFQDEGSGYIMSEDRDHGTHI-YRLADDYLTIVDVACERATDYPYGLESPTII-KKDGLYWFGSGLTSWTDNDNKYSTATDLHGP-WSEWKLFAPEGAKTYDSQVDIVPLDDDPYNS-EHFLFIGDRWQEHDLGNSPIVQMPSIADGVA-SLTWSDTYEGTTHR

bifidum S17 (*BbGal43A*) at 1.4 Å resolution. *BbGal43A* shares 29% sequence identity with the exo- β -1,3-galactanase *Ctl3Gal43A* and also 30% sequence identity with domain I of the *Bacteroides thetaiotaomicron* glycosyl hydrolase from the GH43 family (BT2959; PDB entry 3nqh; Joint Center for Structural Genomics, unpublished work). *BbGal43A* is a single-domain molecule, while both of the available structural homologues are catalytic domains covalently linked to a carbohydrate-binding domain 13 (CBM13; Jiang *et al.*, 2012). *BbGal43A* also shows activity towards synthetic galactosides and evidence of quaternary assembly. Clearly, further structural and biochemical studies are needed to understand the differences between the structural architectures and the enzymatic specificities of these homologues.

2. Materials and methods

2.1. Cloning and purification of *BbGal43A*

The gene encoding *BbGal43A* (GenBank EFR51240.1) from *B. bifidum* strain S17 (NCIMB 41171) was amplified from a genomic DNA library and cloned into the expression vector pET_TrxA by ligation-independent cloning (LIC; Aslanidis & de Jong, 1990) using Phusion High-Fidelity DNA Polymerase (NEB, USA). Both the 912 bp gene product and the linearized vector were treated with 0.33 units of T4 DNA Polymerase (Fermentas, Canada) in the presence of 1× T4 Polymerase Buffer, 4.0 mM DTT and 2.5 mM dATP (for the gene) or 2.5 mM dTTP (for the vector). The reactions were incubated for 30 min at 296 K and were subsequently heat-inactivated for 20 min at 348 K. A 1.0 µl volume of vector treated with T4 Polymerase was mixed with 3.0 µl PCR fragment treated with T4 Polymerase and incubated at 298 K for 30 min, followed by chemical transformation into *Escherichia coli* DH10B and Rosetta(DE3) cells. Positive colonies were selected by colony PCR. Details of primers and source information can be found in Table 1.

Rosetta(DE3) cells containing the recombinant vector were aerobically cultured at 310 K in auto-induction broth (3.3 g l⁻¹ ammonium sulfate, 6.8 g l⁻¹ KH₂PO₄, 7.1 g l⁻¹ Na₂HPO₄, 5.0 g l⁻¹ glycerol, 0.5 g l⁻¹ glucose, 2.0 g l⁻¹ lactose, 10 g l⁻¹ tryptone and 5.0 g l⁻¹ yeast extract) supplemented with

Table 2

Crystallization.

Method	Vapour diffusion
Plate type	Intelli-Plate 96 (Hampton Research)
Temperature (K)	290
Protein concentration (mg ml ⁻¹)	80
Buffer composition of protein solution	0.02 M HEPES pH 7.5, 0.15 M NaCl
Composition of reservoir solution	0.05 M calcium chloride dihydrate, 0.1 M bis-tris pH 6.5, 30% (v/v) polyethylene glycol monomethyl ether 550
Volume and ratio of drop	1.0 µl, 1:1
Volume of reservoir (µl)	80

50 µg ml⁻¹ kanamycin and 33 µg ml⁻¹ chloramphenicol (final concentrations). The cells were induced with 1 mM IPTG at an OD of 0.7 and growth was continued at the decreased temperature of 274 K for 20 h at 200 rev min⁻¹. The cells were harvested by centrifugation (4000g for 20 min at 277 K) and resuspended in 50 ml buffer A (50 mM Tris pH 7.5, 50 mM NaCl, 5.0 mM imidazole, 10% glycerol, 1.0 mM DTT, 0.1 mM PMSF, 200 µg ml⁻¹ lysozyme) per litre of culture. The lysate was then sonicated in an ice bath using a Model 550 Sonic Dismembrator (Fisher Scientific). Cell debris was removed by centrifugation (22 000g for 20 min at 277 K) and the soluble fraction was loaded onto Ni-NTA Superflow Affinity Resin (GE Healthcare) equilibrated with 50 mM Tris pH 7.5, 500 mM NaCl, 10 mM imidazole and eluted using an imidazole gradient (30–500 mM). The fraction containing the purified protein was diluted four times using 50 mM Tris pH 7.5. The Trx-His tag fusion was cleaved with *Tobacco etch virus* (TEV) protease at 275 K overnight. The cleaved protein was loaded onto Ni-NTA Superflow Affinity Resin (GE Healthcare) equilibrated with 50 mM Tris pH 8.0, 10 mM imidazole and then eluted without imidazole.

Size-exclusion chromatography was then performed on a Superdex 200 HiLoad 16/60 column (GE Healthcare) attached to an ÄKTApurifier (GE Healthcare) chromatographic system with 20 mM HEPES pH 7.5, 150 mM NaCl at a flow rate of 0.8 ml min⁻¹. The column was calibrated with a Gel Filtration HMW Calibration Kit (GE Healthcare) according to the manufacturer's instructions. The elution volume and the theoretical mass of *BbGal43A* calculated with *ProtParam* (Wilkins *et al.*, 1999) were compared to estimate the oligomerization state of the molecule in solution. The purity of all samples was estimated by 15% SDS-PAGE.

2.2. Crystallization and X-ray diffraction data collection

The protein was concentrated to 80 mg ml⁻¹ using 10 000 molecular-weight cutoff centrifugal filter concentrators (Millipore, USA) at 4000g and its concentration was determined using a NanoDrop 000 (Thermo Scientific, USA) with a theoretical extinction coefficient of 78 060 M⁻¹ cm⁻¹ at 280 nm (Wilkins *et al.*, 1999). Initial crystallization screening was performed at 291 K using a HoneyBee system (Digilab) and commercial crystallization kits. Details of crystallization can be found in Table 2.

2.3. Data collection and processing

Crystals were harvested, cryoprotected by adding 10%(v/v) ethylene glycol and data were collected on beamline X4C at the National Synchrotron Light Source (NSLS), USA equipped with a MAR 165 CCD detector. The diffraction data were integrated with *XDS* (Kabsch, 2010) and scaled with *AIMLESS* (Evans & Murshudov, 2013). Details of data-collection and processing statistics can be found in Table 3.

2.4. Structure solution and refinement

The solvent content was estimated using the Matthews coefficient (Matthews, 1968) as determined in *phenix.xtriage* (Adams *et al.*, 2010). The protein structure was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using domain I of the *B. thetaiotaomicron* glycosyl hydrolase as the

search model (BT2959; PDB entry 3nqh, 30% sequence identity; Joint Center for Structural Genomics, unpublished work). The model was refined with *phenix.refine* and validated with *MolProbity* and *Coot* (Emsley *et al.*, 2010; Chen *et al.*, 2010). *PyMOL* was used to produce the figures (Schrödinger). Details of the refinement statistics are given in Table 4. Alignment and sequence analyses were performed with *ClustalW* v.2.0 and *ESPrpt* v.3.0 (Robert & Gouet, 2014).

3. Results and discussion

3.1. Production and crystallization of *BbGal43A*

BbGal43A was successfully purified using the protocol described above, with a total protein yield of 40 mg per litre of culture. SDS–PAGE shows that protein is highly pure and its

Figure 1 General features of *BbGal43A*. (a) Chromatogram of *BbGal43A* purification on a HiLoad Superdex 200 16/60 column and linear regression of its calibration with molecular standards. (b) The putative catalytic residues Asp138 and Glu189. (c) An amino-acid sequence alignment shows conservation of the catalytic residues between the homologous structures *BbGal43A*, *Ct1.3Gal43A* and BT2959 (3NQH). Conserved residues are shown in red boxes (conserved in all) and blue boxes (conserved in at least two of the three). β -Strands are denoted as arrows and β -turns are denoted TT. Alignment was performed with *ClustalW2* and the figures were produced with *OriginPro* v.9.0, *PyMOL* and *ESPrpt* v.3.0.

Table 3

Data collection and processing.

Values in parentheses are for the highest resolution shell.

Diffraction source	Beamline X4C, NSLS
Wavelength (Å)	1.1
Temperature (K)	100
Detector	MAR CCD
Space group	$P2_12_12_1$
a, b, c (Å)	64.93, 67.34, 202.13
α, β, γ (°)	90, 90, 90
Resolution range (Å)	47.60–1.40 (1.42–1.40)
Total No. of reflections	1396640 (66108)
No. of unique reflections	174550 (8500)
Mosaicity (°)	0.23
Completeness (%)	99.9 (99.8)
Multiplicity	8.0 (7.8)
$\langle I/\sigma(I) \rangle$	16.6 (1.9)
$R_{p.i.m.}$	0.036 (0.59)
$CC_{1/2}$	0.99 (0.62)
Overall B factor from Wilson plot (Å ²)	14.67

molecular weight is close to the 33 kDa marker (data not shown). The theoretical mass of the sequence, as computed with *ProtParam*, is approximately 34 kDa. Analytical gel filtration was used to estimate the molecular weight of *BbGal43A* in solution. The protein peak eluted at 80.7 ml, suggesting a molecular weight of approximately 67.3 kDa (Fig. 1*a*). This result suggests that the protein is a dimer in solution.

Precipitation assays with 30%(w/v) PEG 4000 and 2.4 *M* ammonium sulfate indicated that protein is highly soluble, since precipitation only took place when the protein concen-

Table 4

Structure-refinement statistics.

Values in parentheses are for the highest resolution shell.

Resolution range (Å)	19.9–1.42 (1.42–1.40)
Completeness (%)	100.0 (99.8)
No. of reflections, working set	174374
No. of reflections, test set	1698
Estimated twinning fraction	0.060 for $k, h, -l$
Final R_{cryst}	0.162
Final R_{free}	0.183
No. of non-H atoms	
Protein	4869
Ligand	14
Water	1261
Total	6144
R.m.s. deviations	
Bonds (Å)	0.011
Angles (°)	1.088
Average B factors (Å ²)	
Protein	17.7
Ligand	52.4
Water	39.3
Ramachandran plot	
Favoured regions (%)	97.3
Additionally allowed (%)	2.5

tration was increased to about 100 mg ml^{−1}. Because of this, we used a concentration of 80 mg ml^{−1} in crystallization trials. Crystals of *BbGal43A* appeared in many conditions from the Index HT screen (Hampton Research), and crystals of 1.0 × 0.4 × 0.4 mm in size grew using 0.05 *M* calcium chloride dehydrate, 0.1 *M* bis-tris pH 6.5, 30%(v/v) polyethylene glycol monomethyl ether 550.

3.2. Structure determination and refinement of *BbGal43A*

BbGal43A crystallized in space group $P2_12_12_1$, with unit-cell parameters $a = 64.9$, $b = 67.3$, $c = 202.1$ Å. The data set was integrated from 20 to 1.4 Å resolution, with 99.8% completeness in the highest resolution shell. The scaled data resulted in a total of 174 550 unique reflections, 1% (1698) of which were set aside as the R_{free} set. The data-collection statistics are presented in Table 3. The Matthews coefficient indicated the presence of between two (solvent content 0.62, $p = 0.38$) and three (solvent content 0.43, $p = 0.57$) molecules in the asymmetric unit. Molecular replacement was performed with *Phaser* (McCoy *et al.*, 2007) and two molecules were found in the asymmetric unit (TFZ = 18.4, LLG = 270.6). Initial model building was carried out with *AutoBuild* in *PHENIX* (Adams *et al.*, 2010) and the final models were hand-built using *Coot* (Emsley *et al.*, 2010). *L*-tests suggested pseudo-merohedral twinning with a twin fraction of 0.06 and operator $k, h, -l$, and this operator was applied in the subsequent refinement steps. The electron density allowed us to model a total of 6144 non-H atoms, of which 6130 were refined against isotropic B factors, as suggested by *phenix.refine* (Afonine *et al.*, 2012) and *PDB_REDO* (Joosten *et al.*, 2014). All refinement was carried out with *phenix.refine* and the model was validated in *MolProbity* (Chen *et al.*, 2010). Translation/libration/screw (TLS) restraints improved the electron density remarkably in the final refinement cycles. At the end of the refinement, 98.5% of the rotamers were in favoured positions, while none

Figure 2

(*a*) Asymmetric unit assembly of the *BbGal43A* crystal structure. This dimer assembly was also suggested by *PISA* to be the quaternary conformation of *BbGal43A* in solution. (*b*) Superposition of *BbGal43A* (green) and *Ct1,3Gal43A* (red)

were in found in a poor position. Also, no residues were found in disallowed regions of the Ramachandran plot, while 97.5% of residues were in favoured regions. A total of 20 residues were assigned in multiple conformations and were refined with split occupancy in both forms. The final model has R_{work} and R_{free} values of 0.16 and 0.18, respectively, and full refinement and validation statistics are given in Table 4.

3.3. Overall structure and quaternary assembly of *BbGal43A*

BbGal43A has a typical fivefold β -propeller fold, with the propellers formed by four β -sheets (Fig. 2*a*). *BbGal43A* has 304 residues, and is significantly smaller than the homologous *Ct1,3Gal43A* (571 residues). *BbGal43A* is remarkably similar to the catalytic domain of the homologous *Ct1,3Gal43A* and BT2959, with r.m.s.d.s of 1.04 and 0.95 Å (both for 302 C α atoms), respectively. The main structural difference between *BbGal43A* and *Ct1,3Gal43A* is the lack of the CBM13 domain (Fig. 2*b*), which was proposed to have a fundamental region for delivery of the substrate from this domain to the active site (Jiang *et al.*, 2012). *BbGal43A* is a single-domain active putative enzyme and may utilize a different, as yet unknown mechanism for substrate allocation.

Structural and sequence homology between *BbGal43A*, *Ct1,3Gal43A* and BT2959 indicate that all three enzymes might be different types of exo- β -1,3-galactanase, although further biochemical characterization needs to be conducted to confirm this. GH43 family members are inverting hydrolases, in which the main proton donor and acceptor residues are supposed to be an aspartate and a glutamic acid, respectively

(Lombard *et al.*, 2014). By comparing the structure with the characterized and homologous *Ct1,3Gal43A*, we suggested the catalytic residues to be Glu112 (auxiliary), Asp138 (nucleophile) and Glu189 (proton donor). The corresponding residues are conserved in putative exo- β -1,3-galactanases homologous to *Ct1,3Gal43A* (Fig. 1*c*).

PISA (Krissinel & Henrick, 2007) analysis suggests that the dimer observed in the asymmetric unit of the determined crystal structure (with a total of two salt bridges and 22 hydrogen bonds between the two molecules) may be similar to its quaternary assembly in solution, and the main interactions are formed between the β 9 strands of each chain (Fig. 3*a*). According to *PISA*, 9.3% of the total surface of the protein (928 Å²) is involved in interactions with the neighbouring chain, where 28 residues forms strong and weak interactions between chains. The most notable are residues 125–132, contained in strand β 9, which contribute 17 hydrogen bonds and one salt bridge to the neighbouring β 9 strand or the loop comprised of residues 69–73 (Fig. 3*a*). As mentioned above, analytical gel filtration also suggests a dimeric configuration of *BbGal43A* in solution (Fig. 1*a*). The formation of a dimer was not observed in homologues of *BbGal43A* such as *Ct1,3Gal43A* (Jiang *et al.*, 2012), and appears to be a unique characteristic of this single-domain group of GH43s. A more complete characterization is required in order to understand the role of dimerization in the activity of *BbGal43A*.

3.4. A bis-tris molecule is bound to the putative active site

The $F_o - F_c$ electron density suggests the presence of a ligand in the putative active site (Fig. 3*b*), which we identified

as a molecule of 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol (BTB) originating from the crystallization conditions (Fig. 3*b*). The electron density was observed in both chains, but we were only able to model BTB in chain *B*. The electron density of the BTB molecule in chain *A* is poor and appears to accept multiple superposed conformations of the molecule. The modelled molecule forms hydrogen bonds to the putative catalytic residue Asp138 (2.7 Å, OD2–O1) and other residues from the putative active site, such as Glu189 (2.6 Å, OE2–O4). It has been proposed that GH43 enzymes require a third catalytic glutamic acid (Ichinose *et al.*, 2005), but the corresponding position is occupied by a glycine in *BbGal43A* (Gly22) and *Ct1,3Gal43A* (Gly54) (Jiang *et al.*, 2012). Furthermore, it was demonstrated that mutation of Glu112 in *Ct1,3Gal43A* completely eliminated the enzyme activity (Jiang *et al.*, 2012). *BbGal43A*, on the other hand, contains Glu82, which could be the third catalytic residue of the enzyme (Fig. 1*b*). More detailed studies using natural enzyme ligands need to be performed in order to understand the architecture of the active site.

4. Conclusions

BbGal43A was successfully cloned, produced, purified and crystallized. The well shaped crystals allowed us to collect X-ray diffraction data to 1.4 Å resolution at the synchrotron source. The structure was successfully solved by molecular replacement and contains two molecules in the asymmetric unit, which might correspond to a quaternary assembly of *BbGal43A* in solution. *BbGal43A* has a typical fivefold β -propeller fold, but lacks the CBM13 domain found in homologous enzymes. A more detailed enzymatic characterization is required to fully understand the biochemical properties of *BbGal43A*.

Acknowledgements

We would like to thank Professor Professor Christian Riedel from Ulm University, Germany for donating *B. bifidum* S17, as well as the RapiData 2013 organization committee. We would also like to acknowledge the grants from the Brazilian funding

agencies Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; Nos. 2008/56255-9, 2009/52840-7, 2009/05328-9, 2011/05712-3 and 2011/20505-4) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Nos. 490022/2009-0, 301981/2011-6 and 400045/2012-5), CAPES and Universidade de São Paulo 'Centro de Instrumentação para Estudos Avançados de Materiais Nanoestruturados e Biosistemas' and 'Núcleo de Apoio à Pesquisa em Bioenergia e Sustentabilidade (NAPBS)'.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Aslanidis, C. & de Jong, P. J. (1990). *Nucleic Acids Res.* **18**, 6069–6074.
- 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.* **D66**, 12–21.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
- Ichinose, H., Yoshida, M., Kotake, T., Kuno, A., Igarashi, K., Tsumuraya, Y., Samejima, M., Hirabayashi, J., Kobayashi, H. & Kaneko, S. (2005). *J. Biol. Chem.* **280**, 25820–25829.
- Jiang, D., Fan, J., Wang, X., Zhao, Y., Huang, B., Liu, J. & Zhang, X. C. (2012). *J. Struct. Biol.* **180**, 447–457.
- Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. (2014). *IUCrJ*, **1**, 213–220.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res.* **42**, D490–D495.
- 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.
- Mohanram, S., Amat, D., Choudhary, J., Arora, A. & Nain, L. (2013). *Sustain. Chem. Process.* **1**, 15.
- Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J. & Beckham, G. T. (2015). *Chem. Rev.* **115**, 1308–1448.
- Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
- Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. & Hochstrasser, D. F. (1999). *Methods Mol. Biol.* **112**, 531–552.