Computational and Structural Biotechnology Journal 19 (2021) 1108–1118

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Plant cell wall architecture guided design of CBM3-GH11 chimeras with enhanced xylanase activity using a tandem repeat left-handed β-3-prism scaffold

Matheus P. Pinheiro [a], Renata A.G. Reis [b], Paul Dupree [c], Richard J. Ward [a,*]

[a] Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP CEP 14040-901, Brazil
[b] Departamento de Física e Química, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, Ribeirão Preto, SP CEP 14040-901, Brazil
[c] Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom

ARTICLE INFO

Article history:
Received 20 October 2020
Received in revised form 7 January 2021
Accepted 8 January 2021
Available online 16 January 2021

Keywords:
Endoxylanase
Carbohydrate-binding module
Protein engineering
Protein design
Multifunctional protein
Lignocellulose hydrolysis
Synthetic biology

ABSTRACT

Effective use of plant biomass as an abundant and renewable feedstock for biofuel production and biorefinery requires efficient enzymatic mobilization of cell wall polymers. Knowledge of plant cell wall composition and architecture has been exploited to develop novel multifunctional enzymes with improved activity against lignocellulose, where a left-handed β–3-prism synthetic scaffold (BeSS) was designed for insertion of multiple protein domains at the prism vertices. This allowed construction of a series of chimeras fusing variable numbers of a GH11 β-endo-1,4-xylanase and the CipA-CBM3 with defined distances and constrained relative orientations between catalytic domains. The cellulose binding and endoxylanase activities of all chimeras were maintained. Activity against lignocellulose substrates revealed a rapid 1.6- to 3-fold increase in total reducing saccharide release and increased levels of all major oligosaccharides as measured by polysaccharide analysis using carbohydrate gel electrophoresis (PACE). A construct with CBM3 and GH11 domains inserted in the same prism vertex showed highest activity, demonstrating interdomain geometry rather than number of catalytic sites is important for optimized chimera design. These results confirm that the BeSS concept is robust and can be successfully applied to the construction of multifunctional chimeras, which expands the possibilities for knowledge-based protein design.

1. Introduction.

Naturally occurring multifunctional proteins (MFPs) are single polypeptide chains comprised of multiple domains, each of which possesses an autonomous function [21]. The strategy of exchanging domains in existing MFPs or fusing novel combinations of individual proteins to form synthetic MFPs displaying a wide range of designed functions has been used to combine individual enzymes for the development of multifunctional biocatalysts [45] and catalytic nanomachines [18]. Modular MFP assembly can also be achieved by creating fusion proteins in which a single polypeptide chain is designed to include all the desired functional domains, typically separated by linker sequences [10]. Although this strategy offers the advantages of synthesis of a single polypeptide and absolute control of domain stoichiometry at the protein expression level, empirical attempts to create novel MFPs by protein fusion can be hampered by poor expression, low solubility and proteolysis [9]. Novel proteins can be designed and constructed by combining protein fragments of from known structures, and the stability of these constructs can be evaluated *in silico* prior to the synthesis and testing phases [25]. Ideally a designed fusion protein should be stable against precipitation and proteolysis, show high protein yields in a common host for heterologous expression, and maintain these robust properties after fusion with added protein domains in the final construct.

Many glycoside hydrolases (GH) found in nature are MFPs, in which one or more catalytic domains are fused to non-catalytic Carbohydrate Binding Modules (CBMs) that present a wide variety of polysaccharide binding specificities [5,17]. Overall, 60% of the more abundant bacterial glycosyl hydrolases (GH) are associated with CBMs, although wide variation is found between different GH families where only 7% of GH8 enzymes are associated with a

* Corresponding author at: Departamento de Química, Bloco 18, Sala 8, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, SP CEP 14040-901, Brazil.
E-mail address: rjward@ffclrp.usp.br (R.J. Ward).

https://doi.org/10.1016/j.csbj.2021.01.011

CBM as compared to 97% of GH9 enzymes [30]. In fungi, less than 13% of the more abundant cellulases include a CBM, although variation is again observed between individual GH families [53]. The functions of CBMs include not only the influence on catalytic properties of hydrolytic enzymes by substrate targeting and substrate proximity [53], but also the amorphogenic perturbation of crystalline substrates resulting in an increased susceptibility to enzymatic attack [2]. As components of MFPs, CBMs may also demonstrate adhesion effects in which the CBM binds to one cell wall component facilitating the activity of an associated catalytic domain against a substrate in close proximity [50]. This functional versatility of CBMs has led to diverse protein engineering applications, including the design and characterization of multienzyme CBM-GH fusions with enhanced catalytic performance for polysaccharide hydrolysis [54].

Full design control over catalytic MFPs requires further information input derived from the nature of the substrate. In the case of the complex biomass derived lignocellulose, this requires detailed knowledge of the architecture of the plant cell wall. Analytic techniques have improved the understanding of the polysaccharide composition of plant cell wall [36], however the conformation of these components and their intermolecular interactions is generally less advanced. Molecular modelling of polysaccharide interactions can provide insights for experimental testing, and a recent example is the interaction of xylan with crystalline cellulose, where composition analysis shows that every second xylose in the xylan backbone is acetylated [6]. Subsequent solid-state NMR studies shows that the xylan polymer adopts a twofold helical screw ribbon that can bind to cellulose microfibrils in the plant cell wall [49] and computer modelling has demonstrated that the xylan in this configuration can intercalate with the hydrophilic face of the crystalline cellulose fibril [6]. We reasoned that the predicted physical proximity created by xylan-cellulose interaction could be tested by creating fusions between the cellulose specific CBM3 and the xylan specific GH11.

The GH11 endoxylanases have been extensively studied due to their potential for biotechnological applications, such as in the pulp and paper industries and for lignocellulose saccharification in biofuels production [39]. Although many synthetic MFPs fusing GH11s with catalytic domains have been constructed and tested (see Table 1, [43]), few designed CBM-GH11 fusion proteins have been reported [29,56]. Using a knowledge-based approach from models of xylan packing on cellulose microfibrils to guide the construction of synthetic MFPs for lignocellulose mapping and degradation, here we describe the design, synthesis and expression of a stable and versatile scaffold protein based on tandem repeats of the left-handed β–3-prism structural motif, which we have named as the Beta Solenoid Scaffold (BeSS). A series of CBM3-GH11-BeSS chimeras were constructed and evaluated in which the interdomain distance, relative orientation and stoichiometry between the CBM and GH11 domains was varied by domain insertion at the vertices of the protein. We demonstrate that several of these CBM3-GH11-BeSS chimeras demonstrated an increased xylo-oligosaccharide release against lignocellulosic substrates derived from common biomass sources, which provides further experimental evidence supporting the proximity of the xylan and crystalline cellulose in these plant cell walls.

2. Materials and Methods.

2.1. Design of a left-handed β–3-prism scaffold – beta solenoid scaffold (BeSS)

Left-hand β-3-solenoid proteins are built from a series of single turns, each comprised of three tandem hexapeptide repeat (HPR) motifs [41]. An initial three-dimensional model of the designed left-handed β–3-solenoid was built from the atomic coordinates of the first 8 HPRs (residues T6 to K53) of the UDP-N-acetylglucosamine acyltransferase from *Leptospira interrogans* (PDB code 3HSQ, [42]). This 8 hexapeptide repeat template was replicated five times using the Coot [13] and Modeller 9.12 [55] software to form a contiguous left-handed β-3-solenoid containing 40 hexapeptide repeats designated as the first version of the Beta Solenoid Scaffold (BeSSv1.0). Proteins with the left-handed β-3-solenoid topology are grouped in the CATH database superfamily 2.160.10.10 [37], which currently includes 25 proteins with known 3D-structures in seven structural clusters from different functional families (Table S1). The atomic coordinates of the β-3-solenoid domains from each of these 25 structures were used to identify the HPR sequences (Table S2), which were analysed to determine amino acid frequencies at each position in the repeat (Table S3). With the aim of improving the stability of the designed β-3-solenoid scaffold, the conformation of the residues at the prism vertices, the hydrophobic core packing and side-chain stacking interactions between neighbouring repeats were examined, and positions 1, 2, 4 and 6 in all 40 HPRs in BeSSv1.0 were replaced by proline, asparagine, isoleucine and glycine residues, respectively, to form the consensus repeat sequence PNxIxG. Three-dimensional modelling of this optimized scaffold with Modeller 9.12 [55]used the structure of BeSSv1.0 as a template, and the optimized structure was designated as BeSSv2.0.

The amino acids at positions 1 (proline) and 6 (glycine) of the consensus HPR are located at the vertices of the solenoid and four of these vertices, corresponding with the connection points of the 8-HPR template, were modified to serve as insertion points for the creation of multidomain protein chimeras. The locations of these insertion sites permit the controlled variation of the interdomain distance, which varies according to the size of the domain and the position of the active site, and the relative orientation (120, 240 and 360°) between the inserted domains. The first insertion point was introduced at amino acids G52 and S53 and was defined by the codon pair GGATCC, the recognition site for the restriction enzyme *Bam*HI. Similarly, the second, third and fourth insertions points were between residues T100/G101 (codons ACCGGT, restriction site for *Age*I), T148/S149 (codons ACTAGT, restriction site for *Spe*I) and G196/T197 (codons GGTACC, restriction site for *Kpn*I), respectively. These modifications containing four unique sites for domain insertion were designated as BeSSv1.1 (modification of BeSSv1.0) and BeSSv2.1 (modification of BeSSv2.0). The modelled structure of BeSSv1.1 and BeSSv2.1 were submitted to molecular dynamics (MD) simulations for stereochemical and stability analysis.

All MD simulations were performed with GROMACS 4.5.x [40] using the AMBER ff99SB forcefield [28] and the TIP3P water model [22] in a cubic simulation box with periodic boundary conditions. Each MD simulation was run for 100 ns at 309 K and an ionic strength of 150 mM NaCl. The isothermal-isobaric (NPT) ensemble was maintained by the v-rescale thermostat [8] during thermalization and by the Nosé–Hoover thermostat [20,35] during production runs. All interaction cut-off values were set to 1 nm and long-range electrostatic interactions were treated by the particle-mesh Ewald (PME) method [15]. The models of BeSSv1.1 and BeSSv2.1 were used as the starting states for independent simulations.

2.2. Construction, expression and purification of BeSSv2.1 and BeSS-CBM3-GH11 chimeras

A series of chimeric multifunctional proteins were created by inserting the glycosyl hydrolase family 11 endo-β-1,4-xylanase from *Bacillus subtilis* (UniProt: P18429; GenBank: NP_389765.1,

PDB 1XXN, GH11) and the cellulose binding domain 3a of the cellulosome scaffoldin CipA from *Clostridium thermocellum* (GenBank: CCV01464.1, PDB 4JO5, CBM3) into the BeSSv2.1. The N- and C-termini of both proteins are in close proximity (Fig. S1A-B) and a series of five CBM3-GH11-BeSS chimeras were constructed, in which the CBM3 domain (C) was introduced at insertion point 1 of the BeSSv2.1, and the GH11 domain (X) at one or more of the remaining three insertion points in the BeSSv2.1. Chimeras with the C at insertion point 1 and a single X at positions 2, 3 or 4 were denominated as CXOO, COXO and COOX, respectively. The chimeras with the C domain at position 1 and multiple X domains at positions 2 and 3, or X domain at positions 2, 3, and 4 were denominated as CXXO and CXXX, respectively. In order to evaluate the physical distances between the active site of the domains GH11 xylanase and CBM3 into the solenoid scaffold BeSSv2.1, the three-dimensional structures of all BeSS-CBM3-GH11 chimeras were modelled using the Modeller 9.12 program [55].

The BeSSv2.1 DNA coding sequence was codon-optimized, synthesized, and cloned into the pET-28a(+) expression plasmid by GenScript (Piscataway, NJ, USA) (Table S4). A DNA fragment encoding CBM3 corresponding to residues 362–531 of the full-length CipA protein was codon-optimized for expression in *Escherichia coli*, synthesized, and cloned into the *Bam*HI (GGATCC) restriction site of the pUC57 vector (GenScript, Piscataway, NJ, USA). Cloning of the fragment generated by *Bam*HI digestion into the corresponding site at the insertion point 1 of the BeSSv2.1 resulted in the insertion of the CBM3 domain yielding the pET28-COOO construct. This construct was used to create vectors pET28-CXOO, pET28-COXO, pET28-COOX, pET28-CXXO, and pET28-CXXX by inserting the GH11 coding sequence into BeSSv2.1 insertion sites 2, 3 and 4 using PCR-generated fragments containing the restriction sites for each insertion point (oligonucleotide sequences shown in Table S5).

The parental GH11 and all BeSS-CBM3-GH11 chimeras were expressed as histidine-tagged fusion proteins after induction with 0.1 mM isopropyl β-D-l-thiogalactopyranoside of transformed *E. coli* (Rosetta) cells in Luria-Broth (LB) medium with antibiotics (50 μg/mL kanamycin and 34 μg/mL chloramphenicol) at 37 °C in an orbital shaker (180 rpm). Cells were harvested by centrifugation (5000*g*, 20 mins) and after lysis by sonication in buffer (50 mM sodium phosphate, pH 8, 300 mM NaCl, 10 mM imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF), 10% (v/v) glycerol, 0.1% (w/v) Triton X-100 and 1X SIGMAFAST™ Protease Inhibitor Cocktail), cell debris was cleared by centrifugation (12000*g*, 30 mins). The proteins were purified from the soluble fraction by immobilized metal affinity chromatography (IMAC) followed by dialysis against 50 mM Tris pH 7.5, 150 mM NaCl for imidazole removal and concentration using membrane filtration (Amicon Ultra-15, Merck KGaA, Darmstadt, Germany).

Protein purity was checked by polyacrylamide gel electrophoresis (PAGE) with Coomassie Brilliant Blue staining. Protein concentration was estimated by the absorbance at 280 nm based on molar extinction coefficients of tryptophan, tyrosine, and cystine residues using a NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA). The purified protein was analyzed by dynamic light scattering (DLS) at a protein concentration of 1.5 mg/mL using a Zetasizer Nano (Malvern Pananalytical Inc, Malvern, UK), and by size exclusion chromatography (SEC) using a Superdex 200 10/300 GL column (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA) in 50 mM MES buffer, pH 5.5, containing 150 mM NaCl.

2.3. Biochemical characterization of the BeSS-CBM3-GH11 chimeras

The specific activity of the GH11 parental enzyme and BeSS-CBM3-GH11 chimeras was measured using wheat arabinoxylan (WAX, Megazyme, Bray, Co Wicklow, Ireland) at a concentration of 5 mg/mL in 0.1 M ammonium acetate pH 6.0. The reactions were equilibrated at 50 °C for 5 mins prior to the addition of purified enzymes at concentrations were adjusted for the number of catalytic domains present in the chimera (2.4 μM for GH11, CXOO, COXO, and COOX; 1.2 μM for CXXO; and 0.8 μM for CXXX), and incubated for an additional 10 mins at 50 °C during which the hydrolysis reaction occurred. Total reducing saccharide release was assayed by addition of 3,5-dinitrosalicylic acid (DNS) to a final concentration of 10 mM and heating at 100 °C for 10 mins. After cooling the reaction for 5 mins on ice the absorbance was measured at 540 nm. The concentration of total saccharide released was calculated using a D-xylose standard curve (concentration range 0.625 to 30 mM). All measurements were repeated in triplicate and expressed as the mean ± standard deviation.

The affinity of BeSS-CBM3-GH11 chimeras for microcrystalline cellulose (Avicel®, Sigma-Aldrich, San Luis, MI, USA) was estimated by incubation of 10 μM of the proteins COOO, CXOO, COXO, COOX, CXXO and CXXX with 1 mg of Avicel® in a total volume of 400 μL of 0.1 M ammonium acetate pH 6.0. The mixtures were incubated at 50 °C for 1 h with orbital agitation (200 rpm) using an Eppendorf thermomixer (Eppendorf AG, Hamburg, Germany) and centrifuged at 14,000*g* for 10 mins. Unbound protein in the supernatants was quantified and the amount subtracted from the total protein offered to estimate the percentage of the protein bound to the crystalline cellulose.

2.4. Activities of BeSS-CBM3-GH11 chimeras against lignocellulosic substrates

Samples of sugarcane bagasse, stem of maize and *Miscanthus* were treated with a ball mixer mill (MM200, Retsch GmbH, Haan, Germany) using 6 cycles of 5 mins grinding at 20 Hz followed by 5 min rest in 65% (v/v) ethanol to produce a homogenized suspension of alcohol insoluble residue (AIR) with a particle size of ~10 μm. The solid material was collected by centrifugation (4000*g*, 10 min) and washed with 100% (v/v) ethanol, then washed with chloroform:methanol (3:2 (v/v)) (under agitation for 3 hrs), then washed with chloroform:methanol (3:2 (v/v)) (under agitation for 12 hrs), followed by successive washes with 100% (v/v), 65% (v/v), 80% (v/v) and 10% (v/v) aqueous ethanol. The pellet was then air dried at 40 °C for three days.

The AIR samples of natural substrates (500 μg) were treated with 4 M NaOH (20 μL) for 1 hr at room temperature (approx. 22 °C) and the pH adjusted 5–6 with 1 M HCl. Assays were performed in 0.1 M ammonium acetate pH 6.0, and enzyme concentration was adjusted according to the number of catalytic domains present in each chimera, as described in the previous section. The reactions (in 1.5 mL volumes) were performed in an Eppendorf thermomixer for 10 min at 50 °C (cycles of 1 min shaking at 500 rpm and 4 mins rest), after which the mixtures were heated at 100 °C for 30 mins to stop the reaction. Aliquots (100 μL) were separated to measure the reducing saccharide release using the DNS assay as described in Section 2.3, and measurements were repeated in triplicate and expressed as the mean ± standard deviation. An aliquot (300 μL) was dried using a centrifugal vacuum evaporator for polysaccharide analysis using carbohydrate gel electrophoresis (PACE).

The PACE experiments were as previously described [16], where chemical derivatization was performed with dried oligosaccharides from 100 μL of enzymatic hydrolysis reaction to which was added 20 μL of a modification solution containing 0.1 M 8-Amino naphthalene-1,3,6-trisulfonic acid (ANTS, Thermofisher, Waltham, MA. USA), 0.1 M 2-picoline-borane (2-PB, Sigma-Aldrich, San Luis, MI, USA) in a 20:17:3 solvent mixture of DMSO: water: glacial acetic acid. After incubation at 37 °C for 16 hrs, the labelling reaction was lyophilized and the derivatized oligosaccharides were

resuspended in 80 μL of 1:7 mixture of Tris-base borate pH 8.2: 6 M Urea. Samples (2 μL) were applied to 20% polyacrylamide gels, and electrophoresis was performed at 200 V for 30 mins and then at 1000 V for 100 min at 10 °C. After electrophoresis, gels were scanned using a G:BOX Chemi HR16 imaging system (Syngene, Cambridge, UK).

3. Results

3.1. Design, expression and characterization of BeSS

Left-handed β-3-solenoid proteins form an elongated triangular prism, where each face is formed from an extended parallel β-sheet comprised of a repeated hexapeptide β-strand [41]. The β-strand conformation forces the side chains of residues at positions 2, 4 and 6 to face the core of the solenoid, whilst the side chains of residues at positions 1, 3 and 5 are on the outer face of the solenoid and interact with the solvent. Each turn of the solenoid is therefore defined by three sequential hexapeptide repeats that form a three β-strand (β/β/β) motif, and consecutive turns are stabilized by cooperative interactions involving main chain hydrogen bonds and hydrophobic interactions of side chains in the core of the prism [23]. Analysis of a total of 448 hexapeptide repeats (Table S2) from the 25 crystal structures of proteins containing left-handed β-3-solenoid domains in the CATH database superfamily 2.160.10.10 (Table S1) revealed the amino acid frequencies for each of the six positions in the repeat (Fig. 1A; Table S3). Structural analyzes of β-3-solenoid proteins were then applied to identify key residues that contribute to increase the stability of the BeSS.

The pair of residues at position 1 in one hexapeptide and position 6 in the previous hexapeptide form the vertices of the prism, and in β-solenoid proteins extra domains are generally inserted between these residues (Fig. S1C). As the main objective was to develop a synthetic polypeptide solenoid scaffold that allows the insertion of catalytic domains of interest in the vertices, the identification of the most frequent amino acids at positions 6 and 1, both in the presence and absence of loop or domain insertions, indicated the amino acids that may be included in these positions both in the presence and absence of the insertions without affecting the stability of the β-solenoid vertices (Table S3). In general, the analysis of the frequency of amino acids in the HPR shows that the most frequent amino acid in both positions 1 and 6 is glycine (Fig. 1A; Table S3), with a ratio of 22 and 13% in positions 1 and 6, respectively. This can be explained by the highly mobile glycine that facilitates the formation of the turns at the vertices of the solenoid. However, analysis of the frequency of amino acids in position 1 in the presence of domains inserted before this hexapeptide repeat revealed a high frequency (20%) of proline residues (Table S3), which may confer rigidity at the start of a new turn and contribute to the stabilization of the β-3-solenoid in the presence of a protein domain inserted in the vertices. Based on these observations, the proline/glycine pair at positions 1 and 6 of the hexapeptide were maintained in the BeSSv2.1 (Fig. 1B). The final BeSSv2.1 design included 4 unique insertion points for the creation of multidomain protein chimera (Fig. 1C), thus the codon pair for the amino acids at positions 1 and 6 at these points were selected to create the unique restriction sites that were used for protein fusion (see materials and methods section for further details).

Analyses of the HPR sequences reveals that only hydrophobic residues occupy position 4, confirming the importance of this position to maintain the hydrophobic core of the solenoid structure, and the most common amino acid at this position, isoleucine (Fig. 1A, Table S3), was assigned to this position in the BeSSv2.1 resulting in an optimized hydrophobic core packing (Fig. 1B). Combined analyses of HPR sequences and β-3-solenoid crystal structures were also used to optimize main chain and side chain interactions between adjacent beta-strands of the β-solenoid scaffold. The side chain of position 2 faces the interior of the solenoid, and amino acid frequency analysis of the HPR sequences showed that residues with different physico-chemical characteristics can occupy this position, including asparagine (Fig. 1A; Table S3). A detailed analysis of the crystal structures of UDP-N-acetylglucosamine acyltransferase from *E. coli* (*Ec*LpxA, PDB code 1LXA) revealed that the main-chain O and N atoms of asparagine at position 2 can form hydrogen bonds with the main-chain N atom of the residue at position 3 and the main-chain O atom of the residue at position 6 of the previous HPR, respectively (Fig. S1D). Furthermore, an asparagine at position 2 can form a continuous string of side-chain interactions in which the ND2 and OD1 atoms hydrogen bond with the OD1 and ND2 of the asparagines at position 2 of the HPR sequences in the previous and subsequent strands of the β-sheets that define the three faces of the prism (Fig. S1D). Based on these analyses, an asparagine residue was assigned to position 2 in the hexapeptides of the BeSSv2.1 (Fig. 1B), which contributes to increasing the number of hydrogen bonded interactions that maintain the BeSSv2.1 stability (Fig. S2B). The side chains of the residues at positions 3 and 5 in the HPR are solvent exposed and frequency analysis reveals high diversity at these positions (Fig. 1A; Table S3). With the aim of maintaining the solubility of the BeSSv2.1, these positions were assigned with residues that are present in the homologous positions of left-hand β-3-solenoid proteins (Fig. 1D). The combined analyses of HPR amino acid frequency and 3D-structure examination therefore led to the definition of the consensus sequence PNxIxG in the BeSSv2.1.

After defining the amino acid sequences of BeSSv1.1 and BeSSv2.1, three-dimensional model structures were generated and analysed by MD simulation. The β-3-solenoid domain of the BeSSv1.1 acquired a pronounced curvature during the simulation, whereas the BeSSv2.1 remained essentially unchanged (Fig. S2A) and is indicative of stronger interactions between the β-strands of the solenoid. This is confirmed by the 1.6-fold higher number of hydrogen bonds and the significantly lower variation in RMSD values in the BeSSv2.1 over the 100 ns simulation time scale (Fig. S2B-C). The reduced stability of the BeSSv1.1 is also seen in the RMSF analysis that show a higher fluctuation of all amino acids along the length the solenoid, and in particular at the N and C-terminal regions (Fig. S2D). Since these results strongly suggested that the BeSSv2.1 was more stable than the BeSSv1.1, and further modelling efforts were therefore focussed on the BeSSv2.1.

The 3D structures of the BeSS-CBM3-GH11 chimeras were also modeled to evaluate the relative distance and orientation between the cellulose binding region of the CBM3 domain at BeSSv2.1 insertion site 1 and the active site regions of the GH11 xylanase domains at insertion sites 2, 3 and 4 (Fig. 2). The cellulose binding site of the CBM3 domain (C), defined by residues N75, W113, D115, H116, Y126, Q169, R171 and W177 [51] and the active site of the GH11 domain (X) inserted at site 2, formed by the residues Y69, E78, Y80, E172, Y174 [34], are separated by a distance of 62 Å and at a relative orientation of 120° (Fig. 2), and the corresponding distances and orientation for the X domains at positions 3 and 4 are 79 Å/240° and 49 Å/0°, respectively (Fig. 2). Natural left-handed beta-solenoid proteins typically present extensive and structurally diverse insertions at the vertices of the beta-solenoid, yet maintain the global left-handed beta-solenoid fold. This strongly suggests that domain insertion does not disrupt the overall protein fold in these proteins. The extensive hydrogen bonding and hydrophobic interactions result in high stability and rigidity of the BeSS, and this leads us to suppose that although the insertion of protein domains will result in structural alterations at the insertion sites, the local-

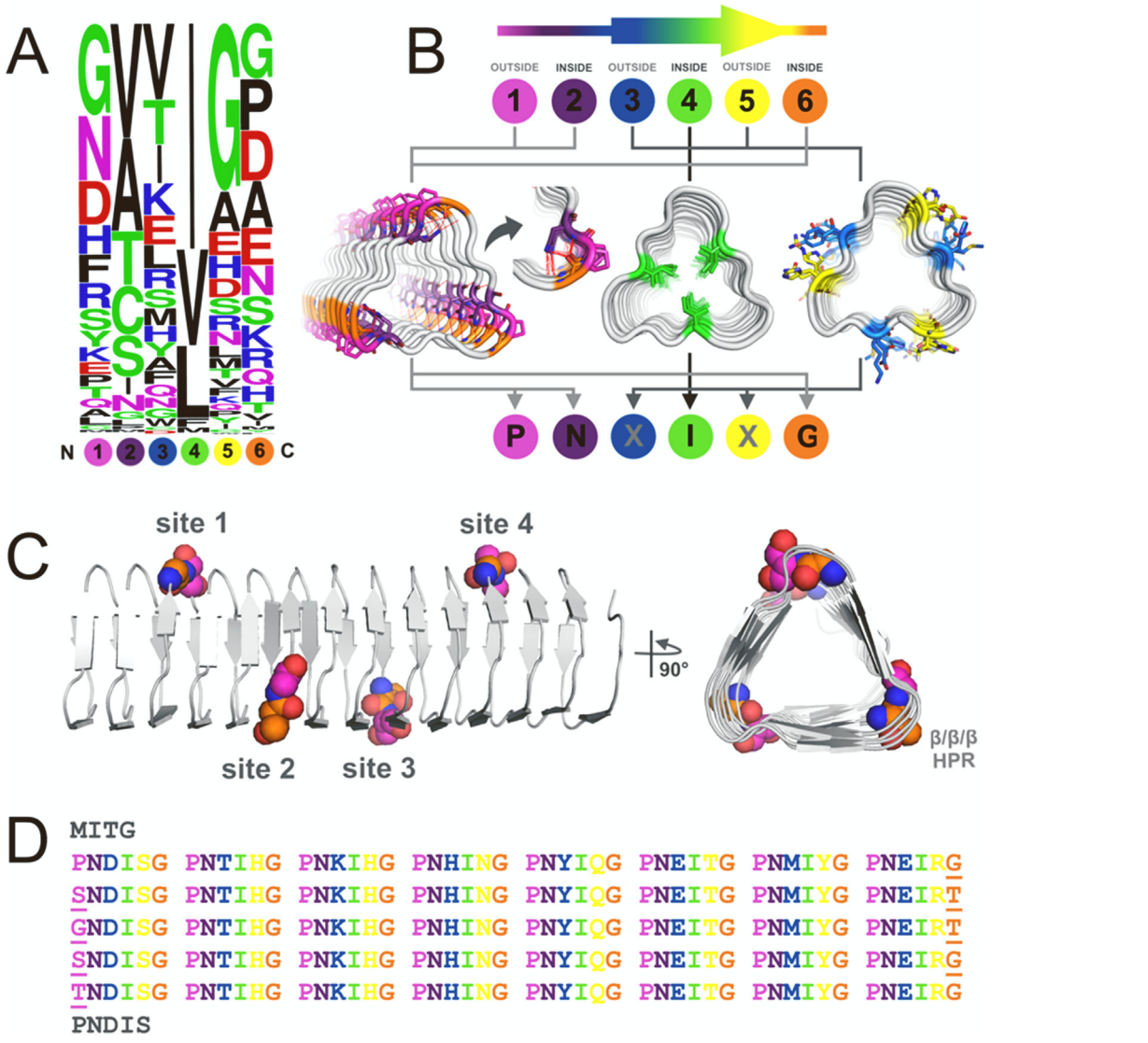

Fig. 1. *(A)* The logo plot of residue frequency for the six positions in the HPR repeats from the 3D structures of proteins with the left-handed β-3-prism structural motif. The plot was generated by WebLogo [11]. *(B)* Representation of the hexapeptide repeat motif of BeSSv2.1 protein. Prolines (position 1) and glycines (position 6) of the hexapeptides are represented in pink and orange, respectively. Asparagine residues (position 2) are colored in purple, isoleucine (position 3) in green, and the residues at position 3 and 5 are colored in blue and yellow, respectively. *(C)* Two orthogonal views of the three-dimensional structure of the BeSSv2.1 modeled structure. The four domain insertion positions at the vertices and the amino acid pairs for insertion of the catalytic domains are represented as spheres. *(D)* The amino acid sequence of the BeSSv2.1 colored according to position in the hexapeptide repeat: 1 (pink), 2 (purple); 3 (blue); 4 (green); 5 (yellow) and 6 (orange). The amino acids encoded by the four unique restriction sites are shown underlined. The GlySer pair at site 1 are encoded by the recognition site for the restriction enzyme BamHI (GGATCC), the ThrGly pair at site 2, the ThrSer pair at site 3 and the GlyThr pair at site 4 are encoded by the recognition sites for restriction enzymes AgeI (ACCGGT), SpeI (ACTAGT) and KpnI (GGTACC), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ized disruption of these prism vertices will not result in the global loss of the of the left-handed beta-solenoid fold.

3.2. Construction, expression and characterization of BeSS-CBM3-GH11 chimeras

The synthetic codon optimized BeSSv2.1 nucleotide coding sequence (Table S4) was expressed in *E. coli* Rosetta as a soluble protein and was purified to homogeneity by IMAC to a final yield of 40 mg/L of culture (Fig. S3A). Analyses by size exclusion chromatography (SEC – Fig. S3B) and dynamic light scattering (DLS – Fig. S3C) are consistent with a homogeneous population of the purified BeSSv2.1 protein with an estimated hydrodynamic radius of 3.1 ± 0.2 nm. Calculation of the radius of gyration from the atomic coordinates of the BeSSv2.1 model using the HYDROPRO software [38] yields a value of 2.75 nm and, assuming a uniform surface hydration layer of 0.145 nm, gives an estimate for the calculated hydrodynamic radius of ~2.9 nm that is in good agreement with the experimental result. The SEC analysis reveals that the major protein peak elutes at a volume of 14.34 mL, which corresponds to a protein with a Stokes radius (R_s) of 4.0 nm as estimated after column calibration with proteins of known R_s. The value for the calculated equivalent radius (R_o) derived from the BeSSv2.1 amino acid sequence using the SEDNTERP1 program [26] is 2.0 nm, which assumes a perfectly spherical protein with a partial specific volume of 0.716 cm^3/g. The geometrical asymmetry of a protein can be estimated from the R_s/R_o ratio [14], which for the BeSSv2.1 gives a value of 2.0 that is in reasonable agreement with the aspect ratio of 2.5 calculated from the model structure of the monomeric protein. The elongated structure of the BeSSv2.1 monomer therefore provides an explanation for the observed elution volume in the SEC analysis.

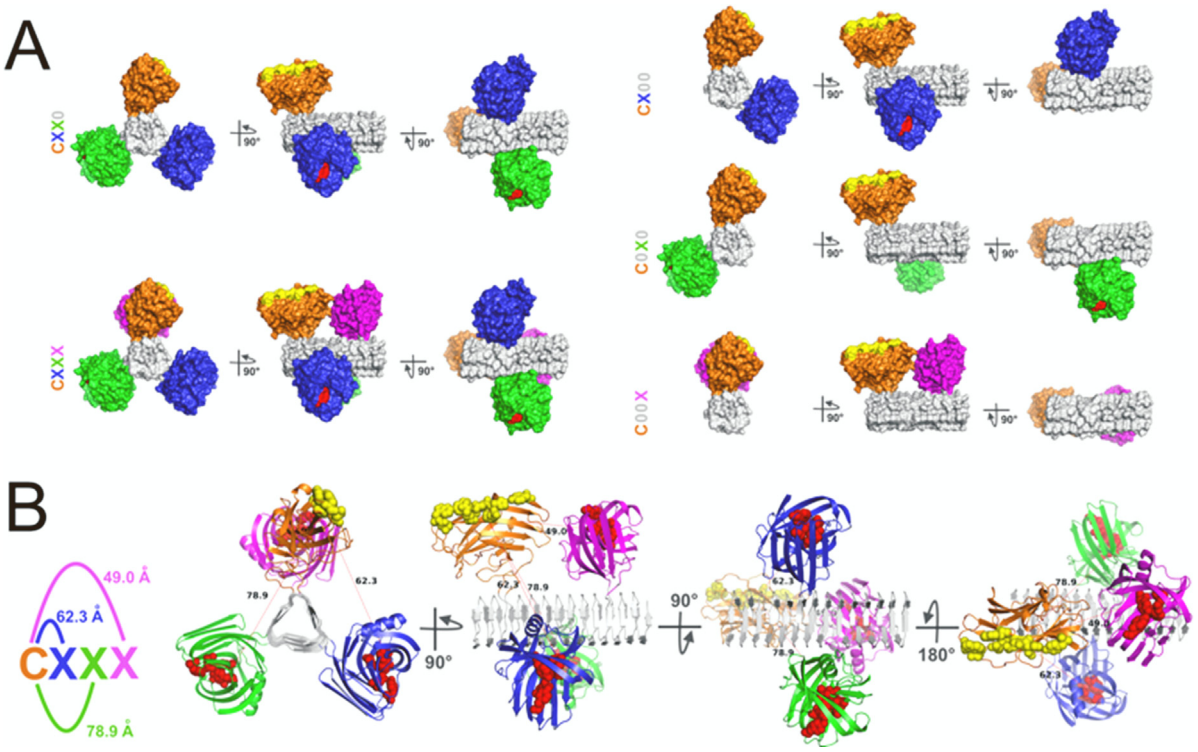

Fig. 2. *(A)* Three surface representation views of the BeSS-CBM3-GH11 chimeras, with the BeSSv2.1 shown in gray. The CBM3 domain (C) at insertion position 1 is shown in orange and the amino acids in the cellulose binding region are highlighted in yellow. The GH11 domains (X) at insertion positions 2, 3 and 4 are colored in blue, green and violet, respectively, with the active site region highlighted in red. *(B)* Four ribbon representation views of the BeSS-CBM3-GH11 with all sites occupied (CXXX). The distances between the CBM cellulose binding region (yellow spheres) and the active sites of the GH11 domains (red spheres) are indicated by red dashes and black text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

All the BeSS-CBM3-GH11 chimeras were successfully constructed, expressed in *E. coli* Rosetta and purified by IMAC (Fig. S4). The endo-xylanase activity of the purified GH11 enzyme and chimeras was evaluated by the release of total reducing saccharides against wheat arabinoxylan (WAX) as substrate, and all chimeras incorporating the GH11 domain presented xylanase activity (Fig. 3A). The chimeras containing single GH11 domains (CXOO, COXO and COOX) and the double GH11 domain chimera (CXXO) present no significant alteration in activity in comparison with the unfused GH11 enzyme. The CXXX chimera shows slightly reduced activity relative to the parental GH11, and since the single GH11 insertions show no difference we conclude that this effect is not the result of inactivation of the enzyme due to the insertion into the BeSS. The activity assays were performed with equivalent molar concentrations of GH11 catalytic domains, such that the activity assay of the CXXX chimera was performed using a third of the molar concentration of protein in comparison with the unfused GH11 enzyme. We speculate that the lower protein concentration reduces the probability of a productive encounter between one of the catalytic domains in the CXXX chimera and the soluble polymeric substrate, thereby lowering the observed measured activity. All the chimeras incorporate the CBM3, and binding to crystalline cellulose of all constructs was assayed by the pull-down assay (Fig. 3B). In this assay, the maintained affinity of the CBM3 domain for crystalline cellulose after insertion into the BeSSv2.1 was validated by measuring the free protein in solution after incubation with Avicel, showing that between 80 and 90% of the chimeras remain associated to the crystalline cellulose.

Following the confirmation of maintained function of the inserted domains, the hydrolytic activity of all BeSS-CBM3-GH11 chimeras was evaluated against the lignocellulose substrates: sugarcane bagasse (Fig. 4A), maize stem (Fig. 4B) and *Miscanthus*

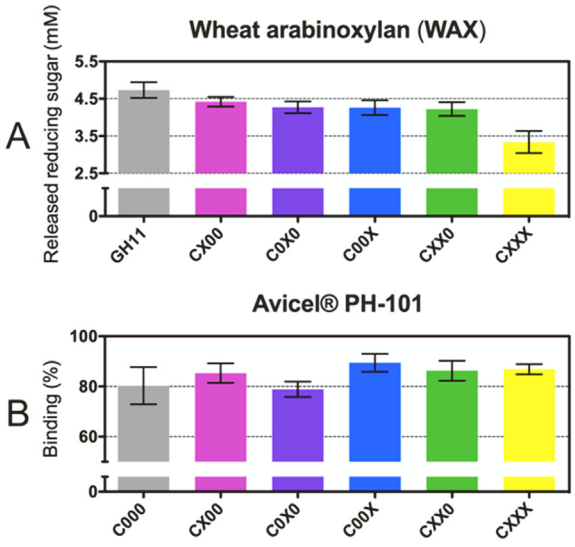

Fig. 3. *(A)* Reducing sugar release by the GH11 and BeSS-CBM3-GH11 chimeras against wheat arabinoxylan (WAX). In order to maintain the concentration of catalytic domains equal in all experiments, the protein concentration for single domain constructs (GH11 alone, and the CXOO, COXO and COOX chimeras) was 2.4 μM, and for the CXXO and CXXX chimeras was 1.2 μM and 0.8 μM, respectively. *(B)* Binding affinity of the BeSS-CBM3-GH11 chimeras and the control (COOO) for crystalline cellulose (Avicel PH-101).

(Fig. 4C). Total reducing saccharide release after 10 mins of reaction was measured by the DNS assay at equimolar concentrations of catalytic domains (Fig. 4), which revealed that under the assay conditions used the CXOO, COXO, COOX and CXXO chimeras pre-

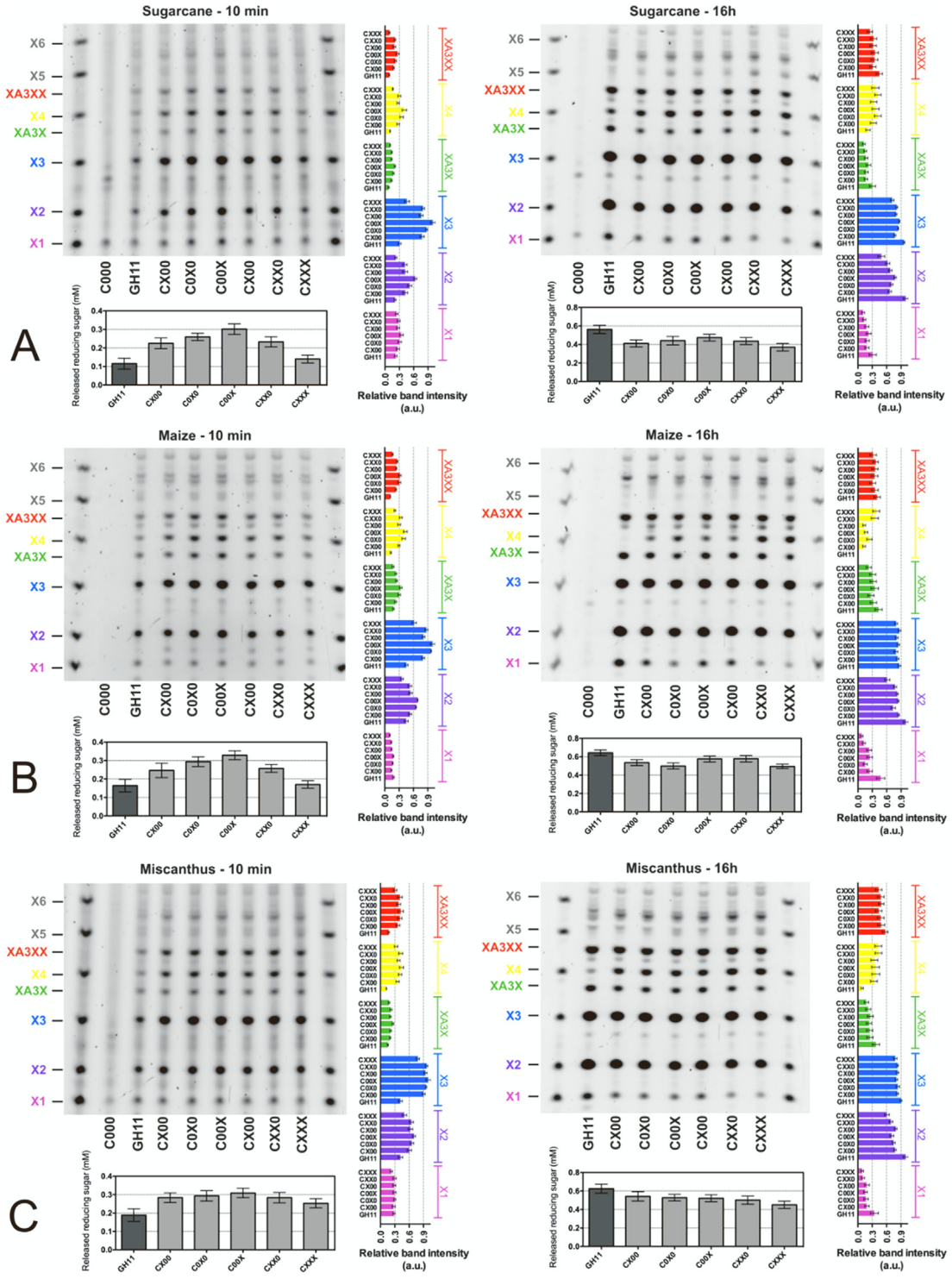

Fig. 4. PACE and total reducing sugar release after 10 min (panels on left) and 16 hr digestion (panels on right) of (A) sugarcane bagasse, (B) maize stems and (C) *Miscanthus* with the GH11 enzyme and the CXOO, COXO, COOX, CXXO and CXXX chimeras. In all panels, the PACE gel show xylose oligomer ladder (X1 to X6, left-hand and right-hand lanes), together with the control (COOO), and the samples from the treated substrates as indicated at the lower edge of the gel. The relative band intensities of selected saccharides X1 (violet), X2 (dark blue), X3 (light blue), XA3X (green), X4 (yellow) and (XA3XX) were analyzed using ImageJ software [46] and presented as histograms on the right-hand side of each PACE gel image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sent a 1.6- to 3.0-fold increase in saccharide release as compared to the non-fused GH11 for sugarcane bagasse and maize substrates, with the COOX chimera being the most active against both substrates (Fig. 4A-B). For these substrates, the CXXX chimera present activities that were not statistically significant as compared with the non-fused GH11 domain. For *Miscanthus*, all BeSS-CBM3-GH11 chimeras present around 1.6-fold increase in saccharide release as compared to the non-fused GH11 (Fig. 4C). However, for all lignocellulose substrates the total reducing saccharide release after the 16 hrs reaction by non-fused GH11 was slightly higher as compared to the BeSS-CBM3-GH11 chimeras, and a comparison of the activity of each chimera shows no statistically signif-

icant difference. This demonstrates that BeSS-CBM3-GH1 chimeras act to rapidly hydrolyze the lignocellulose substrates. For longer reaction times the non-fused GH11 is slightly more efficient than BeSS-CBM3-GH11 chimeras.

Characterization of the released oligosaccharides using PACE after 10 mins treatment of sugarcane bagasse (Fig. 4A), maize stem (Fig. 4B) and *Miscanthus* (Fig. 4C) with the chimeras results in the release of xylose (X1), xylobiose (X2), xylotriose (X3) and xylotetraose (X4), and the detection of faint bands corresponding to xylopentose (X5). In addition to the xylo-oligomers, bands corresponding to xylotriose with an arabinose substituent at the O-3 position of the second xylose (XA3X) and to xylotetraose with an arabinose substituent at the O-3 position of the second xylose (XA3XX) were also detected. These saccharides are consistent with the pattern of oligosaccharide hydrolysis by the GH11 against the arabinoxylan present in the cell wall [4]. Treatment of the lignocellulose substrates for 16 hrs with the BeSS-CBM3-GH11 chimeras yields a similar profile of oligosaccharide release, although all the bands are more intense in the case of the CXOO, COXO, and particularly for the COOX chimera. The increase in the band intensities observed in the PACE experiments correlates with the results of the DNS assay, and it is noteworthy that the intensity of all bands increases in the PACE, and therefore the increase in reducing saccharide release measured in the DNS assay is not due to an artefact arising from changes in length of the oligosaccharide on treatment with the chimeras. The initial rapid hydrolysis observed after treatment with the chimeras followed by the later and higher oligosaccharide levels with the unfused GH11 indicates that the xylan in the tested plant cell walls is present as a relatively exposed fraction associated with the cellulose fibres that is accessible to the chimeras, and a less exposed fraction that is preferentially available to the smaller unfused GH11. This general effect shows slight variation between the different lignocellulose sources that were tested. For example, after 16 hrs treatment of sugarcane bagasse the unfused GH11 released higher levels of X1, X2 and X3 than the chimeras, whereas with maize and *Miscanthus* this difference was less pronounced, probably reflecting variation in availability of the xylan between the different sources of lignocellulose. The exception to this overall pattern is X4, which shows significantly increased levels with all chimeras after both short and 16hrs treatments as compared to the unfused GH11. Previous studies have established that unbranched xylooligosaccharides shorter than X5 are poor substrates for the *Bacillus subtilis* GH11 [12], and the accumulation of X4 after treatment with the chimeras in the current work suggests that the engineered enzymes may increase the liberation of hydrolysis products from the lignocellulose substrate.

4. Discussion

As part of ongoing work to further the understanding of the structural architecture of plant cell wall polysaccharides and the role this plays in the saccharification of biomass derived lignocellulose, we have focused on the development of multifunctional proteins to explore the relative spatial distributions between the xylan chains and the crystalline cellulose chains of plant cell walls. We were inspired by previous studies in which CBM binding to one cell surface component results in the activity of an associated catalytic domain on a neighboring polysaccharide substrate, such as the adhesion effect against fungal cell walls of the CBM35 domain in the CsxA, an exo-β-D-glucoaminidase from the bacteria *Amycolatopsis orientalis* [32], and the adherence to the host cell surface of the CBM71 domain in the BgaA, a β-galactosidase from the pathogen *Streptococcus pneumoniae* [50]. We reasoned that the proximity of plant cell wall components could be inferred by observing the effects of combining a CBM specific for one cell wall component with a catalytic domain that is specific for a neighboring polysaccharide. We were guided by a detailed computer modelling of the interaction of xylan with the (010) face of the crystalline cellulose fibril in *Arabidopsis thaliana* [6], which is based on the knowledge that every second xylose in the xylan backbone is acetylated[6]and supported by solid-state NMR results demonstrating that in the context of the plant cell wall the xylan adopts a twofold (2_1-fold) helical screw conformation that is in register with the cellulose Iβ microfibril [49]. Modification of every second xylose by alternating arabinosyl and glucuronosyl residues together with computer modelling indicates a similar interaction between gymnosperm xylans and the (110) face of the cellulose microfibril [7]. These studies suggest the intercalation of decorated xylans is possible at both of the hydrophilic surfaces of crystalline cellulose fibres. The present study not only aimed to demonstrate how this knowledge of the cell wall architecture can drive the design of novel chimeric hydrolases, but also to develop the BeSS protein scaffold that can be used as a molecular gauge that allows control of the distance and relative orientation between two functional domains. Finally, the hydrolytic capabilities of a series of designed chimeric enzymes based on the BeSSv2.1 fused with the CBM3a of the CipA cellulosomal scaffoldin from *Clostridium thermocellum* and the GH11 β-1,4-endoxylanase from *Bacillus subtilis* were tested against several lignocellulosic substrates derived from common agricultural biomass feedstocks.

The CipA CBM3a binds specifically to crystalline cellulose [33] and an electron microscope study has revealed that the protein associates to the hydrophilic (110) face of cellulose microfibrils [27]. The 3D structure of the CBM3a [57] presents a highly conserved linear polar/aromatic amino acid motif shown to be the CBM3 cellulose binding site by site-directed mutagenesis [3] and by NMR studies of the homologous *B. subtilis* CBM3d-cellopentose complex [44]. The polar/aromatic binding site is located on the edge of a mainly planar surface containing, and is consistent with glycosyl-protein interactions at the (110) face of cellulose microfibrils mediated by hydrogen bonding interactions between polar side chains and exposed glycoside –OH groups and hydrophobic stacking of pyranose rings with aromatic residues on the flat binding surface of the CBM3. Results presented in the present study confirm the CBM3 binding to crystalline cellulose, and demonstrate that this function is maintained in all the chimeric CBM3-GH11-BeSS chimeras.

The GH11 β-1,4-endoxylanases (EC 3.2.1.8) present a highly conserved β-jellyroll architecture [52], where the xylan substrate binds to a surface cleft that includes a conserved aromatic residue cluster and the catalytic residues Glu78 and Glu172. The GH11 enzymes recognize a pattern of three consecutive unsubstituted D-xylopyranose units in glucuronoarabinoxylans and specifically hydrolyze the β-1,4-glycosidic bond one D-xylopyranose unit before a methylglucuronic acid branch [4], and can hydrolyze acetylated xylans although with reduced catalytic efficiency [1]. Here we have demonstrated that the GH11 β-1,4-endoxylanase A from *Bacillus subtilis* (*Bs*XynA) retains full activity against the soluble WAX substrate when fused to the BeSS indicating that the active sites remain exposed for substrate binding, and suggests no proximity effects result from interactions between the catalytic domains and the BeSS. In addition, the profile of oligosaccharides released by the chimeric enzymes against the various biomass substrates is consistent with the known specificity and substrate preferences of GH11 catalysis.

The PACE results demonstrate that the initial release of xylooligomers by all CBM3-GH11-BeSS chimeras is significantly increased as compared with the unfused enzyme. This result is consistent with the computer-generated models of the xylan decorated cellulose microfibril, and we propose that the CBM3 domain directs

binding of the chimeras to the hydrophilic face where the GH11 domain hydrolyzes the associated xylan polymer. These results add to a growing body of evidence suggesting that the interactions between the cell wall polysaccharides can be understood through functional studies comparing the catalytic activities of glycosyl hydrolases and their corresponding CBM-enzyme fusions. The close proximity between cellulose microfibres and xylans has been inferred from the observed increase in xylan removal by a fusion between the CipA CBM3a and the *Neocallimastix patricarium* GH11 xylanase against tobacco stem sections [19], and proximity effects have also been inferred from the increased mannan degradation by a CipA CBM3a-GH5 mannanase fusion in *Physocomitrella* cell walls [58]. The effects of these synthetic constructs offer an explanation for the fusion of catalytic domains with non-cognate CBMs that is observed in many multidomain glycosyl hydrolases found in nature. In these proteins, natural selection appears to favour combinations in which the CBMs localize catalytic domains to different polysaccharides in close proximity in the cell wall matrix rather than direct association with the enzyme substrate.

The majority of studies with protein chimeras have used an end-to-end domain fusion strategy, whereas the present study employs insertional fusion in a synthetic scaffold that allows precise control of relative distance and orientation between the selected domains. The effects of the CXOO, COXO and COOX constructs are greater that with the CXXO and CXXX, and since the amount of enzyme was adjusted to maintain a constant active site concentration, this indicates that the average total activity from multiple sites is lower in comparison to the single site constructs and suggests that the observed effect is the result of the activity in only one catalytic domain. Furthermore, the COOX construct in which the GH11 and CBM3 domains are oriented on the same vertex of the BeSS shows the highest activity, showing that the alignment of the CBM3 binding site and GH11 active site favours hydrolysis. It is noteworthy that the enhanced activities observed in all single catalytic domain constructs suggests that optimal alignments between the CBM3 and GH11 domains may also be achieved by partial rolling of the cellulose bound chimeras around the long axis, although this would introduce internal strain that might explain the lesser effects of the CXOO and COXO construct. The presence of one or more additional GH11 domain in another vertex of the BeSS might reduce the mobility of the bound protein, and contribute to the reduced effects observed in the CXXO and CXXX.

These results demonstrate that different interdomain orientations influence the activity of the CBM3-GH11-BeSS chimeras, and highlight the importance of component geometry in MFP design. Previous studies have shown that the effects of protein fusion can depend on the sequence in which the individual protein components are joined, for example variations in the order of the *Thermobifida fusca* exoglucanase Cel6B fusions with CBM2, CBM3 and L2 domains, together with variations in the interdomain linkers, resulted in differing catalytic efficiencies against insoluble cellulose substrates [47]. These efforts have contributed to the design of synthetic cellulosomes, where the recombination of the same subset of components in different constructs with alternative geometries has a significant effect on catalytic performance [31,47]. These studies, together with the findings in the current work demonstrate how scaffold protein design can control the distance and relative orientation of catalytic domains for the successful design of MFPs. The BeSS allows the rapid evaluation of multiple combinations of catalytic domains in which stoichiometry, interdomain distance and relative orientation can be varied. Furthermore, the creation of MFPs by domain fusion often results in polypeptides with molecular weights > 80 kDa and the heterologous expression of such large recombinant proteins in *E coli* often results in low expression yield, poor solubility and proteolytic breakdown [48]. We have demonstrated that all the BeSS chimeras in the present study were expressed as soluble proteins at pH values where most glycosyl hydrolases are active, are compatible with IMAC purification, and undergo only minimal proteolysis, all of which provides significant practical advantages for MFP construction and testing. These useful properties are likely to be derived from the stable fold and low RMSD of the designed scaffold, possible stabilizing effects of interdomain interfaces between the BeSS and the fused domains, and the use of insertional domain fusion which reduces the length and access of the interdomain linker regions to host cell proteases. Finally, we note that in both the CipA CBM3a and the BsXynA crystal structures the N- and C-termini are within 5–10 Å of each other, which facilitates their fusion in the BeSS insertion sites. Approximately 54% of the proteins in the PDB database show close proximity (<5Å) between their N- and C-termini [24], which indicates that many native proteins may readily be inserted into the BeSS. For proteins with more distant N- and C-termini, fusion with BeSS would require the use of longer peptide linkers or prior engineering by circular permutation.

In conclusion, we have successfully designed the BeSS, and incorporated a CBM3 and variable numbers of the GH11 catalytic domain into the scaffold. The designs proved to be robust when expressed in *E. coli*, and the designed MFP demonstrated enhanced catalytic properties against lignocellulosic substrate. Although our results demonstrate that all insertion sites in the BeSS can be successfully occupied, the interdomain geometry rather than number of catalytic sites is more important for optimized chimera design. The results also validate the knowledge-based approach to chimera design using computer models of the interactions between plant cell wall components, which brings a new dimension to enzyme engineering for biomass deconstruction, and the design strategy could be applied to gain insights as to the distance constraints between cell wall components. Furthermore, domain fusion with BeSS allows the creation of multifunctional polypeptides with Mwts above 100 kDa that are robust during expression and which can be readily purified. We suggest that the BeSS could find applications in other areas where precise distance and orientation of domains and good expression levels are important for MFP function. In addition to lignocellulose degradation, other possible applications include the design of MFPs for therapeutics, multidomain biosensors, and biologically compatible hydrogels for tissue engineering.

Author contributions

M.P.P. performed the experiments; R.J.W., M.P.P and P.D. designed the experiments and analyzed the data; R.A.G.R. performed and analyzed the molecular dynamics simulations; R.J.W. and P.D. conceived the project and coordinated the experimental activities; R.J.W., P.D. and M.P.P prepared and edited the manuscript; and R.J.W. agrees to serve as the author responsible for contact and ensures communication.

Declaration of Competing Interest

The authors declare no conflict no interest associated with the work described in this manuscript.

Acknowledgements

This work was funded by São Paulo Research Foundation Research Grant (FAPESP 2016/24139-6; R.J.W.), and Fellowships (FAPESP 2012/01066-2 and 2014/13956-8 BEPE; M.P.P.), National Council for Scientific and Technological Development Fellowship (CNPq 305788/2017-5; R.J.W.), the National Institute of Science and Technology of Bioetanol (INCT-Bioetanol) (FAPESP 2011/57908-6 and 2014/50884-5, CNPq 574002/2008-1 and 465319/2014-9) (Brazil).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.01.011.

References

[1] Arai T, Biely P, Uhliariková I, Sato N, Makishima S, Mizuno M, Nozaki K, Kaneko S, Amano Y. Structural characterization of hemicellulose released from corn cob in continuous flow type hydrothermal reactor. J Biosci Bioeng 2019;127(2):222–30. https://doi.org/10.1016/j.jbiosc.2018.07.016.

[2] Arantes V, Saddler JN. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. Biotechnol Biofuels 2010;3:4. https://doi.org/10.1186/1754-6834-3-4.

[3] Benhar I, Tamarkin A, Marash L, Berdichevsky Y, Yaron S, Shoham Y, Lamed R, Bayer EA. Phage display of cellulose binding domains for biotechnological application. In: Glycosyl Hydrolases for Biomass Conversion. American Chemical Society; 2000. p. 10–168. https://doi.org/10.1021/bk-2001-0769.ch010.

[4] Biely P, Singh S, Puchart V. Towards enzymatic breakdown of complex plant xylan structures: State of the art. Biotechnol Adv 2016;34(7):1260–74. https://doi.org/10.1016/j.biotechadv.2016.09.001.

[5] Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 2004;382(3):769–81. , http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15214846.

[6] Busse-Wicher M, Gomes TCF, Tryfona T, Nikolovski N, Stott K, Grantham NJ, Bolam DN, Skaf MS, Dupree P. The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of Arabidopsis thaliana. Plant J 2014;79(3):492–506. https://doi.org/10.1111/tpj.12575.

[7] Busse-Wicher M, Li A, Silveira RL, Pereira CS, Tryfona T, Gomes TCF, Skaf MS, Dupree P. Evolution of Xylan Substitution Patterns in Gymnosperms and Angiosperms: Implications for Xylan Interaction with Cellulose. Plant Physiol 2016;171(4):2418–31. https://doi.org/10.1104/pp.16.00539.

[8] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys 2007;126(1).

[9] Chen R, Chen Q, Kim H, Siu K-H, Sun Q, Tsai S-L, Chen W. Biomolecular scaffolds for enhanced signaling and catalytic efficiency. Curr Opin Biotechnol 2014;28:59–68. https://doi.org/10.1016/j.copbio.2013.11.007.

[10] Chen X, Zaro JL, Shen W-C. Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev 2013;65(10):1357–69. https://doi.org/10.1016/j.addr.2012.09.039.

[11] Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res 2004;14(6):1188–90. https://doi.org/10.1101/gr.849004.

[12] Cuyvers S, Dornez E, Moers K, Pollet A, Delcour JA, Courtin CM. Evaluation of the xylan breakdown potential of eight mesophilic endoxylanases. Enzyme Microb Technol 2011;49(3):305–11. https://doi.org/10.1016/j.enzmictec.2011.05.005.

[13] Emsley P, & Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr (2004), 60, 2126–2132. https://doi.org/S0907444904019158 [pii]10.1107/S0907444904019158.

[14] Erickson HP. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. Biol Procedures Online 2009;11:32–51. https://doi.org/10.1007/s12575-009-9008-x.

[15] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald Method. J Chem Phys 1995;103(19):8577–93.

[16] Goubet F, Jackson P, Deery MJ, Dupree P. Polysaccharide analysis using carbohydrate gel electrophoresis. A method to study plant cell wall polysaccharides and polysaccharide hydrolases. Anal Biochem 2002;300(1):53–68. https://doi.org/10.1006/abio.2001.5444.

[17] Guillén D, Sánchez S, Rodríguez-Sanoja R. Carbohydrate-binding domains: multiplicity of biological roles. Appl Microbiol Biotechnol 2010;85(5):1241–9. https://doi.org/10.1007/s00253-009-2331-y.

[18] Hartley CJ, Williams CC, Scoble JA, Churches QI, North A, French NG, Nebl T, Coia G, Warden AC, Simpson G, Frazer AR, Jensen CN, Turner NJ, Scott C. Engineered enzymes that retain and regenerate their cofactors enable continuous-flow biocatalysis. Nature Catal 2019;2(11):1006–15. https://doi.org/10.1038/s41929-019-0353-0.

[19] Hervé C, Rogowski A, Blake AW, Marcus SE, Gilbert HJ, Knox JP. Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. PNAS 2010;107(34):15293–8. https://doi.org/10.1073/pnas.1005732107.

[20] Hoover WG. Canonical dynamics: equilibrium phase-space distributions. Phys Rev A General Phys 1985;31(3):1695–7. https://doi.org/10.1103/PhysRevA.31.1695.

[21] IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). (1989). Nomenclature for multienzymes. European Journal of Biochemistry, 185(3), 482–485. DOI:10.1111/j.1432-1033.1989.tb15140.x_1.

[22] Jorgensen WL. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J Am Chem Soc 1981;103(2):335–40. https://doi.org/10.1021/ja00392a016.

[23] Kobe B, Kajava A. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. Trends Biochem Sci 2000;25(10):509–15.

[24] Krishna MMG, Englander SW. The N-terminal to C-terminal motif in protein folding and function. Proc Natl Acad Sci 2005;102(4):1053–8. https://doi.org/10.1073/pnas.0409114102.

[25] Kuhlman B, Bradley P. Advances in protein structure prediction and design. Nat Rev Mol Cell Biol 2019;20(11):681–97. https://doi.org/10.1038/s41580-019-0163-x.

[26] Laue TM, Shah BD, Ridgeway TM, & Pelletier SL. Analytical Ultracentrifugation in Biochemistry and Polymer Science. In Harding A, Rowe S (eds.). Royal Society of Chemistry. (1992).

[27] Lehtiö J, Sugiyama J, Gustavsson M, Fransson L, Linder M, Teeri TT. The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. PNAS 2003;100(2):484–9. https://doi.org/10.1073/pnas.212651999.

[28] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010;78(8):1950–8. https://doi.org/10.1002/prot.22711.

[29] Liu LW, Cheng J, Chen HG, Li XQ, Wang SY, Song AD, Wang MD, Wang B, Shen JW. Directed evolution of a mesophilic fungal xylanase by fusion of a thermophilic bacterial carbohydrate-binding module. Process Biochem 2011;46(1):395–8.

[30] Mba Medie F, Davies GJ, Drancourt M, Henrissat B. Genome analyses highlight the different biological roles of cellulases. Nat Rev Microbiol 2012;10(3):227–34. https://doi.org/10.1038/nrmicro2729.

[31] Mingardon F, Chanal A, Tardif C, Bayer EA, Fierobe H-P. Exploration of new geometries in cellulosome-like chimeras. Appl Environ Microbiol 2007;73(22):7138–49. https://doi.org/10.1128/AEM.01306-07.

[32] Montanier C, van Bueren AL, Dumon C, Flint JE, Correia MA, Prates JA, Firbank SJ, Lewis RJ, Grondin GG, Ghinet MG, Gloster TM, Herve C, Knox JP, Talbot BG, Turkenburg JP, Kerovuo J, Brzezinski R, Fontes CMGA, Davies GJ Gilbert HJ. Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. Proc Natl Acad Sci USA 2009;106(9):3065–70. https://doi.org/10.1073/pnas.0808972106.

[33] Morag E, Lapidot A, Govorko D, Lamed R, Wilchek M, Bayer EA, Shoham Y. Expression, purification, and characterization of the cellulose-binding domain of the scaffoldin subunit from the cellulosome of Clostridium thermocellum. Appl Environ Microbiol 1995;61(5):1980–6.

[34] Murakami MT, Arni RK, Vieira DS, Degreve L, Ruller R, & Ward RJ. Correlation of temperature induced conformation change with optimum catalytic activity in the recombinant G/11 xylanase A from Bacillus subtilis strain 168 (1A1). FEBS Lett. (2005), 579(28), 6505–6510. https://doi.org/S0014-5793(05)01299-8 [pii]10.1016/j.febslet.2005.10.039.

[35] Nosé S. A molecular dynamics method for simulations in the canonical ensemble. Mol Phys 1984;52(2):255–68. https://doi.org/10.1080/00268978400101201.

[36] O'Neill MA, & York WS. The Composition and Structure of Plant Primary Cell Walls. In Annual Plant Reviews online (pp. 1–54). American Cancer Society; 2018. DOI:10.1002/9781119312994.apr0067.

[37] Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH - a hierarchic classification of protein domain structures. Structure 1997;5(8):1093–108. https://doi.org/10.1016/S0969-2126(97)00260-8.

[38] Ortega A, Amorós D, García de la Torre J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. Biophys J 2011;101(4):892–8. https://doi.org/10.1016/j.bpj.2011.06.046.

[39] Polizeli ML, Rizzatti AC, Monti R, Terenzi HF, Jorge JA, Amorim DS. Xylanases from fungi: properties and industrial applications. Appl Microbiol Biotechnol 2005;67(5):577–91. https://doi.org/10.1007/s00253-005-1904-7.

[40] Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D, Hess B, Lindahl E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 2013;29(7):845–54. https://doi.org/10.1093/bioinformatics/btt055.

[41] Raetz CR, Roderick SL. A left-handed parallel beta helix in the structure of UDP-N-acetylglucosamine acyltransferase. Science 1995;270(5238):997–1000. , http://www.ncbi.nlm.nih.gov/pubmed/7481807.

[42] Robins LI, Williams AH, Raetz CR. Structural basis for the sugar nucleotide and acyl-chain selectivity of Leptospira interrogans LpxA. Biochemistry 2009;48(26):6191–201. https://doi.org/10.1021/bi900629e.

[43] Saadat F. A review on chimeric xylanases: methods and conditions. 3. Biotech 2017;7(1):67. https://doi.org/10.1007/s13205-017-0660-6.

[44] Santos CR, Paiva JH, Sforca ML, Neves JL, Navarro RZ, Cota J, Akao PK, Hoffmam ZB, Meza AN, Smetana JH, Nogueira ML, Polikarpov I, Xavier-Neto J, Squina FM, Ward RJ, Ruller R, Zeri AC, & Murakami MT. Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from Bacillus subtilis 168. Biochem J (2012), 441(1), 95–104. https://doi.org/BJ20110869 [pii] 10.1042/BJ20110869

[45] Schmidt-Dannert C, Lopez-Gallego F. A roadmap for biocatalysis – functional and spatial orchestration of enzyme cascades. Microb Biotechnol 2016;9(5):601–9. https://doi.org/10.1111/1751-7915.12386.

[46] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012;9(7):671–5. https://doi.org/10.1038/nmeth.2089.

[47] Setter-Lamed E, Moraïs S, Stern J, Lamed R, Bayer EA. Modular organization of the Thermobifida fusca exoglucanase Cel6B impacts cellulose hydrolysis and designer cellulosome efficiency. Biotechnol J 2017;12(10):1700205. https://doi.org/10.1002/biot.201700205.
[48] Shukla S, Toleti SR. Heterologous expression and purification of a 238 kDa large biofilm associated surface protein (Bap) in Escherichia coli. Cloning Transgenesis 2016;5:1000153. https://doi.org/10.4172/2168-9849.1000153.
[49] Simmons TJ, Mortimer JC, Bernardinelli OD, Pöppler A-C, Brown SP, DeAzevedo ER, Dupree R, Dupree P. Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. Nature Commun Commun 2016;7(13902):1–9. https://doi.org/10.1038/ncomms13902.
[50] Singh AK, Pluvinage B, Higgins MA, Dalia AB, Woodiga SA, Flynn M, Lloyd AR, Weiser JN, Stubbs KA, Boraston AB, King SJ. Unravelling the multiple functions of the architecturally intricate Streptococcus pneumoniae beta-galactosidase. BgaA. PLoS Pathogens 2014;10(9). https://doi.org/10.1371/journal.ppat.1004364.
[51] Tormo J, Lamed R, Chirino AJ, Morag E, Bayer EA, Shoham Y, Steitz TA. Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. EMBO J 1996;15(21):5739–51. https://doi.org/10.1002/j.1460-2075.1996.tb00960.x.
[52] Torronen A, Harkki A, Rouvinen J. Three-dimensional structure of endo-1,4-beta-xylanase II from Trichoderma reesei: two conformational states in the active site. EMBO J 1994;13(11):2493–501. https://doi.org/10.1002/j.1460-2075.1994.tb06536.x.
[53] Várnai A, Mäkelä MR, Djajadi DT, Rahikainen J, Hatakka A, Viikari L. Carbohydrate-binding modules of fungal cellulases: occurrence in nature, function, and relevance in industrial biomass conversion. Adv Appl Microbiol 2014;88:103–65. https://doi.org/10.1016/B978-0-12-800260-5.00004-8 #.
[54] Walker JA, Takasuka TE, Deng K, Bianchetti CM, Udell HS, Prom BM, Kim H, Adams PD, Northen TR, Fox BG. Multifunctional cellulase catalysis targeted by fusion to different carbohydrate-binding modules. Biotechnol Biofuels 2015;8(1):1–20. https://doi.org/10.1186/s13068-015-0402-0.
[55] Webb B, Sali A. Protein structure modeling with MODELLER. Methods Mol Biol 2017;1654:39–54. https://doi.org/10.1007/978-1-4939-7231-9_4.
[56] Yang A, Cheng J, Liu M, Shangguan Y, Liu L. Sandwich fusion of CBM9_2 to enhance xylanase thermostability and activity. Int J Biol Macromol 2018;117:586–91. https://doi.org/10.1016/j.ijbiomac.2018.05.199.
[57] Yaniv O, Morag E, Borovok I, Bayer EA, Lamed R, Frolow F, Shimon LJ. Structure of a family 3a carbohydrate-binding module from the cellulosomal scaffoldin CipA of Clostridium thermocellum with flanking linkers: implications for cellulosome structure. Acta Crystallogr Sect F Struct Biol Cryst Commun 2013;69(Pt 7):733–7. https://doi.org/10.1107/S174430911301614X.
[58] Zhang X, Rogowski A, Zhao L, Hahn MG, Avci U, Knox JP, Gilbert HJ. Understanding how the complex molecular architecture of mannan-degrading hydrolases contributes to plant cell wall degradation. J Biol Chem 2014;289(4):2002–12. https://doi.org/10.1074/jbc.M113.527770.