

OPEN

# A processive endoglucanase with multi-substrate specificity is characterized from porcine gut microbiota

Weijun Wang<sup>1</sup>, Tania Archbold<sup>1</sup>, Joseph S. Lam<sup>2</sup>, Matthew S. Kimber<sup>2</sup> & Ming Z. Fan<sup>1</sup>

Cellulases play important roles in the dietary fibre digestion in pigs, and have multiple industrial applications. The porcine intestinal microbiota display a unique feature in rapid cellulose digestion. Herein, we have expressed a cellulase gene, *p4818Cel5\_2A*, which singly encoded a catalytic domain belonging to glycoside hydrolase family 5 subfamily 2, and was previously identified from a metagenomic expression library constructed from porcine gut microbiome after feeding grower pigs with a cellulose-supplemented diet. The activity of purified *p4818Cel5\_2A* was maximal at pH 6.0 and 50°C and displayed resistance to trypsin digestion. This enzyme exhibited activities towards a wide variety of plant polysaccharides, including cellulosic substrates of avicel and solka-Floc®, and the hemicelluloses of  $\beta$ -(1 → 4)/(1 → 3)-glucans, xyloglucan, glucomannan and galactomannan. Viscosity, reducing sugar distribution and hydrolysis product analyses further revealed that this enzyme was a processive endo- $\beta$ -(1 → 4)-glucanase capable of hydrolyzing cellulose into cellobiose and cellotriose as the primary end products. These catalytic features of *p4818Cel5\_2A* were further explored in the context of a three-dimensional homology model. Altogether, results of this study report a microbial processive endoglucanase identified from the porcine gut microbiome, and it may be tailored as an efficient biocatalyst candidate for potential industrial applications.

Plant lignocellulosic biomass, mainly including cellulose, hemicelluloses, lignin and pectin, is the largest renewable natural resource for the potential production of biofuel, biomaterials, and chemical feedstocks<sup>1</sup>. This biomass material also constitutes the major portion of dietary fibers that are utilized by various animal species<sup>2</sup>. The anaerobic fermentation of dietary fibers by gut microbiota is critical in maintaining the health of host animals. This fermentation assists in the supply of energy, vitamins and minerals, detoxification of noxious compounds, modulation of the host's immune system, and protection of the host from colonization by pathogenic microorganisms<sup>2-5</sup>. On the other hand, some specific types of dietary fibers, such as cellulose and lignin, have been considered as an anti-nutritional factor particularly for monogastric food production animals<sup>6</sup>. Cellulose is the most abundant plant polysaccharide and the major component of dietary fibers. Cellulases are a group of enzymes that decompose cellulosic material, typically including endoglucanases, exoglucanases and  $\beta$ -glucosidase, and play concerted roles in cellulose digestion in animals<sup>2</sup>. Moreover, cellulases are also widely used in multiple industries<sup>7</sup>, including the food, textile, laundry detergent, pulp and paper, as well as in the livestock production sectors for improving feed conversion efficiency and enteric health<sup>8</sup>. For this reason, tremendous efforts have been directed to the research on cellulolytic enzymes.

Metagenomic mining is a powerful tool in the search for novel cellulose-degrading enzymes, either through enzyme functional screening<sup>9</sup> or through gene cataloguing by high-throughput sequencing<sup>10,11</sup>. Functional screening of expression libraries has the advantage over gene cataloguing approaches in that enzyme activity is assayed directly, and thus allows the discovery of new enzyme functionalities and novel enzyme families with no sequence similarity to previously characterized enzymes<sup>9</sup>. To date, function-based metagenomic screening has unearthed a diversity of cellulases from the gut microbiota of various species<sup>12-14</sup> and environmental samples<sup>15,16</sup>.

<sup>1</sup>Department of Animal Biosciences, University of Guelph, Guelph, Ontario, N1G 2W1, Canada. <sup>2</sup>Department of Cellular and Molecular Biology, University of Guelph, Guelph, Ontario, N1G 2W1, Canada. Correspondence and requests for materials should be addressed to M.Z.F. (email: [mfan@uoguelph.ca](mailto:mfan@uoguelph.ca))

Gut microbiomes encompass a diverse repertoire of cellulase genes<sup>17</sup>. Metagenomic analyses have revealed tremendous sequence diversity for cellulases from the gut microbiomes of humans, rodents, and ruminants<sup>2,18,19</sup>. In contrast, metagenomic analyses and functional characterization of gut microbial cellulases in monogastric food animals are relatively limited. Pigs are known for their ability to readily adapt to a wide variety of high-fiber diets<sup>20</sup>. Digesta passage rate throughout the porcine gut system is relatively rapid<sup>20</sup>, as further compiled in Table S1. Thus, a unique gut microbiota has been established in the pig for dietary fibre degradation<sup>21,22</sup>. Our previous functional metagenomic screen of the pig cecum and colon microbiomes identified 18 carbohydrate-active enzyme (CAZyme) genes in total, eleven of them were co-localized and formed four function-related gene clusters coding enzymes potentially for cellulose,  $\beta$ -(1  $\rightarrow$  4)-D-mannan and pectin degradation<sup>23</sup>. A recent metagenomic analysis of porcine microbiome also detected 29 families of CAZyme sequences likely associated with cellulose, xylan and pectin degradation<sup>24</sup>.

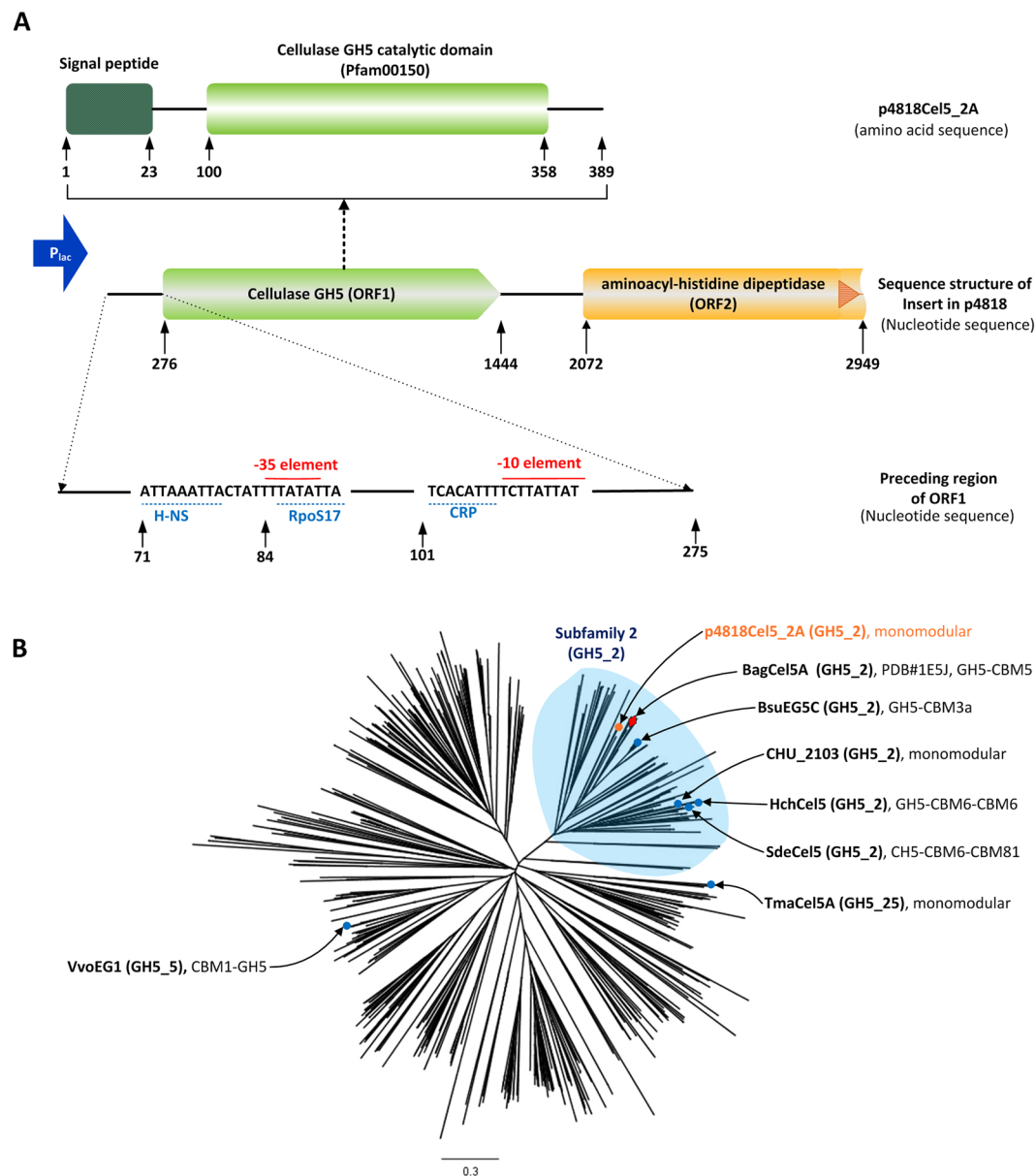
Relative to the vast amount of sequence data generated through metagenomic analyses, biochemical characterization data for these candidate cellulolytic enzymes is scarce. Given the limitation of computational prediction in the detail protein function from sequences, mapping biochemical function to an exponentially increasing number of deposited enzyme sequences represents one of the major challenges in the post-metagenomic analysis era<sup>25</sup>. Indeed, most of the current biochemical information regarding cellulolytic enzymes was obtained from the isolated and pure cultured microorganisms<sup>26,27</sup>. Considering the challenge in culturing most of animal gut microorganisms, the biochemical characterization for the cellulolytic enzymes identified directly from gut microbiome will add novel insights into the molecular mechanism of dietary fiber digestion in animal hindgut. We thus set out to mine novel cellulases from the pig gut microbiome by a metagenomic functional screening. One of the discovered cellulase genes (*p4818Cel5\_2A*) was further expressed for the detail biochemical characterization. The resulting enzyme proved trypsin resistant, and have a broad substrate specificity towards a variety of plant polysaccharides. More specifically, it acted as a monomeric and processive  $\beta$ -(1  $\rightarrow$  4)-endoglucanase in hydrolyzing cellulose into cellobiose and celotriose as the primary end products.

## Results and Discussion

**Diet induction and functional metagenomic screening identified a novel cellulase gene from the pig gut microbiome.** The porcine hindgut microbiota is a unique ecosystem, capable of rapid cellulose degradation (Table S1)<sup>21,22</sup>. A metagenomic plasmid expression library was previously constructed from the pooled hindgut microbiota of six grower pigs, which were enriched for crystalline cellulose degradation by feeding with a diet containing 10% Solka-Floc<sup>®</sup> as the sole fiber source (Table S2) for 28 days, as detailed in our previous study<sup>23</sup>. The functional screening against this library (Fig. S1A,B) had resulted in the identification of 18 CAZyme genes, and eleven of them were co-localized and formed four function-related gene clusters<sup>23</sup> (Fig. S2).

One previously identified positive clone<sup>23</sup>, referred to as p4818 (Fig. S2), was sequenced by primer walking. The DNA insert is 2949-bp long with 39% G + C content (GenBank accession number, MH373350), containing two tandem open reading frames (ORFs). The ORF at 3' side was predicted as a partially C-terminal truncated aminoacyl-histidine dipeptidase with a closest homologue of an aminoacyl-histidine dipeptidase from *Faecalibacterium prausnitzii* KLE1255 (Genbank accession number: ZP07799867, sequence identity of 60%). The other ORF at 5' side was predicted to encode a cellulase with the closest homologue being an uncharacterized endoglucanase from *Eubacterium rectale* DSM 17629 (Genbank accession number: CBK89462, sequence identity of 51%) (Fig. 1A). Both bacterial species belong to the class *Clostridia*. However, through the megablast with the entire insert nucleotide sequence against nucleotide (nr/nt) databases, no sequence hit could be found for having the significant nucleotide sequence similarity to the inquiry, suggesting that this insert sequence was likely derived from an unsequenced *Clostridia* species. Furthermore, we performed an amino acid sequence search for this cellulase against EBI-metagenome database - Mgnify (<https://www.ebi.ac.uk/metagenomics>; search parameters: E < 0.01, gap penalty of 0.02 for open and 0.04 for extend). This resulted in 2416 hits, with the closest homologue being MGYP000014238623 (amino acid sequence identity of 60.3%) from the human digestive sub-datasets in "Host-associated Biome Database (v.2018\_06)". The similar search against animal sub-dataset (v.2018\_06) resulted in 135 hits, with a closest homologue being MGYP000572419700 (amino acid sequence identity of 57.0%). Together, this indicates that homologues of *p4818Cel5\_2A* (a member of GH5\_2, see the result below) are widely present in the gut microbiomes of human and other animal species. Consistently, GH5 family enzyme sequences were found to be abundant among human microbiomes<sup>19</sup>. Immediately upstream of the start codon of this putative cellulase, the potential -35 (TTATA) and -10 (TCTTATTAT) promoter elements were identified. Canonical binding sites for three transcription factors were also found in this region, including CRP (cAMP receptor protein, also known as catabolite activator protein; CAP), H-NS (histone-like nucleoid structuring protein) and RpoS (RNA polymerase,  $\delta^S$  subunit) (Fig. 1A). CRP is a cAMP-activated global transcription factor for catabolite repression particular in microbial carbon and energy metabolism<sup>28</sup>. While, RpoS is a central regulator of the general stress response, and its association with CRP has been found in microbial metabolism regulation<sup>29</sup>. Lastly, H-NS is a component of bacterial chromatin, and influences gene expression at both local and global scale<sup>30</sup>. Together, these observations imply a coordinated regulation in the expression of this novel cellulase gene under pig hindgut environment.

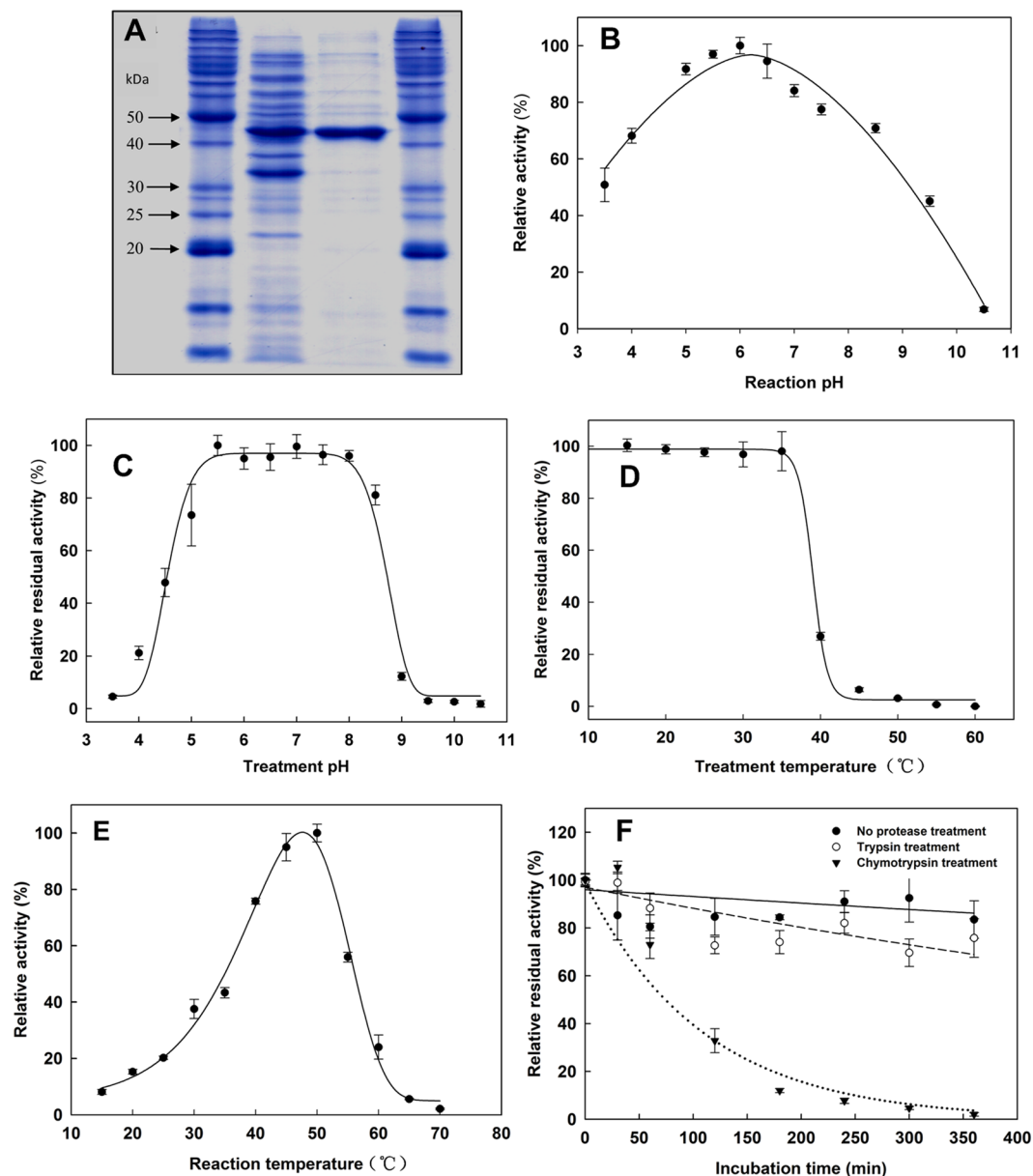
This predicted cellulase ORF comprises 1170 nucleotides, encoding a protein of 389 amino acids (Fig. 1). The presence of an N-terminal signal peptide (1–23 amino acids) is consistent with the expected extracellular location of this enzyme. The cellulase was predicted to be a member of the family 5 glycosyl hydrolases (GH5) (CAZy database, <http://www.cazy.org/>), based on its clustering with known GH5\_2 subfamily sequences in a phylogenetic analysis<sup>31</sup> (Fig. 1B), and thus belongs to subfamily 2 in GH5 family. Accordingly, it was designated as *p4818Cel5\_2A* cellulase. Although the catalytic domain(s) of cellulases are often associated with specific carbohydrate-binding modules (CBMs)<sup>32</sup>, no canonical CBM is recognizable in *p4818Cel5\_2A* sequence, and thus this enzyme is a monomeric cellulase.



**Figure 1.** Organization of the insert from the p4818 positive clone and phylogenetic analysis of p4818Cel5\_2A with other characterized members of the GH5 family. **(A)** Organization of the insert from positive clone of p4818. Two open reading frames (ORFs) were identified within this insert. The predicted  $-10$  and  $-35$  promoter elements are shown in red above the sequences and the identified transcription factor binding sites are underlined in blue. The prediction for bacterial promoters was performed using the online tool of BPROM-Prediction<sup>67</sup> against 274 bp sequence preceding to the cellulase p4818Cel5\_2A coding region. Enzyme modules were identified using the Simple Modular Architecture Research Tool (SMART, <http://smart.embl-heidelberg.de/>). **(B)** Phylogenetic analysis of p4818Cel5\_2A with sequences of characterized GH5 enzymes. All sequences were aligned using MUSCLE in Geneious version 8.0.5. The tree was constructed using Geneious Tree Builder version 8.05. The other reported processive GH5 endoglucanases were indicated by blue dot, and the sequence of template structure (PDB#1E5J) was indicated by red prism.

**Expression, purification and characterization of p4818Cel5\_2A cellulase.** GH5 enzymes display a variety of hydrolytic activities<sup>31</sup>, making sequence-based function predictions challenging for this family. The sequence encoding the mature form of p4818Cel5\_2A (with the 1–23 amino acids signal peptide sequence removed) was inserted into a pET28a expression vector, fused in frame with an N-terminal His-tag. The recombinant p4818Cel5\_2A was purified with a typical yield of 19 mg protein for per liter of *E. coli* culture. SDS-PAGE analysis of the purified protein depicted a protein with an apparent molecular mass of 43 kDa (Fig. 2A), which is consistent with its theoretical molecular mass of 43.4 kDa.

The p4818Cel5\_2A activity on carboxymethyl cellulose (CMC) exhibited a reaction pH optimum at 6.0, retaining >70% of its maximal activity from pH 4.0 to 8.5 (Fig. 2B). It retained over 90% of its original activity



**Figure 2.** Purification and general properties of p4818Cel5\_2A enzyme. (A) Purification of p4818Cel5\_2A, Lanes 1,4: Molecular weight standards; lane 2: Crude extract; and lane 3: Ni-NTA purified p4818Cel5\_2A cellulase. (B) The effect of reaction pH on p4818Cel5\_2A activity on CMC. The activity was presented as relative activity(%) to the specific activity of 1816.0  $\mu\text{mol}/(\text{mg protein} \cdot \text{min})$  measured at optimum pH 6.0. A constant ionic strength buffers containing 100 mM Tris, 50 mM acetic acid and 50 mM MES over a pH range from 4.0 to 10.0 were used. (C) pH stability of the purified p4818Cel5\_2A enzyme activity relative to % of the residual activity after the incubation under pH 7.0. (D) The relative residual activity of purified p4818Cel5\_2A after treatment at the indicated temperature relative to % of the residual activity after the incubation at 15 °C. (E) The effect of reaction temperature on p4818Cel5\_2A activity relative to % of the activity measured at 50 °C. (F) Resistance of p4818Cel5\_2A activity to porcine trypsin and bovine chymotrypsin treatment relative to % of the p4818Cel5\_2A activity measured without any protease treatments. The p4818Cel5\_2A cellulase in the final concentration of 50  $\mu\text{g}/\text{ml}$  was incubated with either 5000 U/ml trypsin (BAEE unit) or 200 U/ml chymotrypsin (BTEE unit) at 37 °C in 100 mM MES buffer (pH 6.0). The purified P4818Cel5\_2A cellulase of 1.6  $\mu\text{g}$  and 1.0% carboxymethyl cellulose (CMC) were used in all assays for (B–E). Values were expressed as means  $\pm$  SE,  $n = 3$ .

after storage for 24 h at 4 °C in the pH range from 5.5 to 8.0 (Fig. 2C). The enzyme retained ~95% and ~30% activity after 30 min incubation under 35 °C and 40 °C, respectively; however, 95% activity was lost at temperatures >45 °C (Fig. 2D). Interestingly, the optimal reaction temperature of p4818Cel5\_2A for 10-min reaction was at 50 °C (Fig. 2E), a temperature where the isolated enzyme was unstable, possibly indicating that the binding of the enzyme to CMC significantly increased the enzyme's thermostability. In addition, the effects of selected divalent metal ions and chemicals on the enzyme activity were also examined (Table S3). The p4818Cel5\_2A activity

towards CMC was inhibited ( $P < 0.05$ ) in the presence of 1.0 mM Ni<sup>2+</sup> (by 12%), Cu<sup>2+</sup> (by 12%), Mn<sup>2+</sup> (by 30%), Cd<sup>2+</sup> (by 34%), and Zn<sup>2+</sup> (by 58%), respectively. The inhibitory effects of these metal ions are likely due to their ability to bind to the active site of p4818Cel5\_2A. In addition, the reducing agent of 1,4-dithiothreitol (DTT) at 5.0 mM demonstrated an activation effect ( $P < 0.05$ ) by 27%. DTT was also previously reported to activate an endoglucanase from the anaerobic bacteria, *Clostridium thermocellum*<sup>33</sup>.

**p4818Cel5\_2A is resistant to trypsin digestion.** Enzymes secreted by microbiota in the lumen environment of cecum and the proximal colon will be exposed to residual exocrine pancreatic proteases. The ability of p4818Cel5\_2A to withstand proteolysis by trypsin and chymotrypsin was therefore examined. p4818Cel5\_2A retained >70% of its original activity on CMC after 6 h of incubation with trypsin at 5,000 U/ml under 37 °C, whereas it lost 95% of its activity after an analogous incubation with chymotrypsin at 200 U/ml (Fig. 2F). In comparison, a GH5 family cellulase from *Bacillus amyloliquefaciens* (BamCel5, Megazyme) lost ~90% of its original activity after a 5 h treatment by either trypsin or chymotrypsin under the same conditions as for p4818Cel5\_2A (Fig. S3). Furthermore, the online protease cutting sites server (PeptideCutter, [https://web.expasy.org/peptide\\_cutter/](https://web.expasy.org/peptide_cutter/)) predicts 37 and 24 cutting sites for p4818Cel5\_2A by chymotrypsin and trypsin, respectively; in contrast, 28 and 27 were found for the catalytic domain of BamCel5. In general, the total number of potential cutting sites will not necessarily correlate with susceptibility to a given protease as many candidate sites are likely buried and therefore not accessible to the protease active site. Empirical identification of sites within p4818Cel5\_2A that are actually cleaved by a given protease, along with structure determination will provide more insights into the molecular basis of its resistance to trypsin. Our observation is consistent with the previous report that porcine cecal digesta retained a high residual trypsin activity, while chymotrypsin activity was reduced to negligible levels<sup>34</sup>. To the best of our knowledge, the resistance of microbial enzymes secreted by the gut symbiotic microflora to exocrine pancreatic proteases has been scarcely reported. This feature reflects an environment-driven enzyme evolution for the resistance to proteolysis. It is of great benefit for the potential application of p4818Cel5\_2A as an exogenous feed enzyme in food animal industry, as it allows more of the activity to reach its targeted site of action.

**p4818Cel5\_2A has a broad substrate specificity.** GH5 family of enzymes exhibited a wide variety of glycoside hydrolytic activities, including endoglucanase, endoxylanase, endomannanase, as well as exoglucanase, exomannanase,  $\beta$ -glucosidase,  $\beta$ -mannosidase and others. It is one of the largest and most diverse GH families, with more than 50 subfamilies<sup>31</sup>. Specifically, members in subfamily 2 of GH5 family (GH5\_2) are often associated with endo- $\beta$ -(1  $\rightarrow$  4)-glucanase activity<sup>31</sup>. Of note, a bifunctional cellulase/chitosanase activity has been found for several members in this subfamily<sup>35</sup>. Herein, the substrate specificity of p4818Cel5\_2A was examined against a variety of polysaccharides (Table 1). The enzyme exhibited the highest activity on  $\beta$ -(1  $\rightarrow$  4)/(1  $\rightarrow$  3)-glucan from barley, followed by CMC, hydroxyethyl cellulose (HEC), glucomannan from Konjac, xyloglucan from tamarind seed, and galactomannan from Locust bean gum. All of these polysaccharides are soluble and contain common  $\beta$ -(1  $\rightarrow$  4)-glucosidic linkages in their backbone with the exception of galactomannan, which has a linear backbone of  $\beta$ -(1  $\rightarrow$  4)-linked D-mannose modified (typically) sub-stoichiometrically by  $\alpha$ -(1  $\rightarrow$  6) linked D-galactose residues. The catalytic efficiency constant ( $V_{\max}/K_m$ ) further verified that barley  $\beta$ -(1  $\rightarrow$  4)/(1  $\rightarrow$  3)-glucan was the best substrate for p4818Cel5\_2A among all the tested polysaccharides (Table 2; Fig. S4). In contrast, no activity was detected for curdlan [ $\beta$ -(1  $\rightarrow$  3)-glucan] from *Alcaligenes faecalis* and laminarin [ $\beta$ -(1  $\rightarrow$  3)-glucan backbone with  $\beta$ -(1  $\rightarrow$  6)-glucosidic substitutions] from *Laminaria digitata*, indicating that p4818Cel5\_2A hydrolyzes  $\beta$ -(1  $\rightarrow$  4)-D-glucosidic bond. Specifically, p4818Cel5\_2A activity on xyloglucan was ~170 times lower than that on  $\beta$ -(1  $\rightarrow$  4)/(1  $\rightarrow$  3)-glucan (Table 1). Xyloglucan consists of a  $\beta$ -(1  $\rightarrow$  4)-linked glucan backbone that is further substituted at O6 with xylosyl residues through  $\alpha$ -(1  $\rightarrow$  6)-linkage, suggesting that the bulky O6-substitution at the glucose unit sterically hinders the enzyme catalysis.

p4818Cel5\_2A was active towards different forms of insoluble cellulosic substrates, where activity towards the regenerated amorphous cellulose (RAC) was higher than the crystalline celluloses of Avicel and Solka-Floc<sup>®</sup> by 6–12 times (Table 1). These differences likely reflect the much lower substrate accessibility of Avicel and Solka-Floc<sup>®</sup> to p4818Cel5\_2A than that of RAC, as the crystallinity index (CI) values of Avicel (CI at 0.5–0.6) and Solka-Floc<sup>®</sup> (CI at 0.4–0.7) are higher than that of RAC (CI at 0 ~ 0.04)<sup>36</sup>. Furthermore, the near 2-fold higher p4818Cel5\_2A activity towards Avicel than Solka-Floc<sup>®</sup> was likely due to the fact that Avicel has a lower degree of polymerization (DP at 150–500 glucose units) compared with Solka-Floc<sup>®</sup> (DP at 750–1500 glucose units)<sup>36</sup>. In addition, p4818Cel5\_2A exhibited a specific activity of 9.0 and 5.3  $\mu\text{mol}$  glucose equivalent  $\cdot \mu\text{mol}^{-1}$  protein  $\cdot \text{min}^{-1}$  on Avicel and Solka-Floc<sup>®</sup>, respectively. These values are 7 to 220 folds higher than the corresponding activities of many other reported cellulases, such as a GH9 cellulase (Cel9B) and a GH51 cellulase (Cel51A) from ruminal *Fibrobacter succinogenes*<sup>37</sup>, a GH48 cellulase (Cel48A) from *Thermobifida fusca*<sup>38</sup>, two GH5\_2 cellulases of Cel5H and Cel5 from *Saccharophagus degradans*<sup>39</sup> and from *Hahella chejuensis*<sup>40</sup>, respectively. Given the similar properties of Solka-Floc<sup>®</sup> to pre-treated cellulosic substrates<sup>36</sup>, the hydrolytic activity of p4818Cel5\_2A towards Avicel and Solka-Floc<sup>®</sup> is thus of great interest in enhancing cellulosic biomass valorization and animal fiber digestion.

Although the enzymes from subfamily 2 in GH5 family (GH5\_2) generally display endo- $\beta$ -(1  $\rightarrow$  4)-glucanase activity<sup>31</sup>, they can differ greatly in their substrate utilization profiles. For example, the activities towards Avicel, RAC,  $\beta$ -(1  $\rightarrow$  4)/(1  $\rightarrow$  3)-glucan and CMC were previously measured for three GH5\_2 endo- $\beta$ -(1  $\rightarrow$  4)-glucanases from *Bacillus subtilis* BS-5<sup>41</sup>, *Volvariella volvacea*<sup>42</sup> and *Hahella chejuensis*<sup>40</sup> respectively. Whereas, GH5\_2 cellulase of CelDZ1 from hot spring microbiota was found only active on CMC and  $\beta$ -D-glucan, but not on Avicel<sup>43</sup>, and another GH5\_2 cellulase from soil was found active on CMC,  $\beta$ -D-glucan and Avicel, while inactive to xyloglucan<sup>44</sup>. In comparison, p4818Cel5\_2A displayed activities towards a wide variety of plant polysaccharides, including regenerated amorphous cellulose, highly crystalline cellulosic substrates (e.g., Avicel and Solka-Floc<sup>®</sup>), as well as the hemicelluloses of  $\beta$ -(1  $\rightarrow$  4)/(1  $\rightarrow$  3)-glucan, xyloglucan, glucomannan and galactomannan (Table 1).

Substrate	Main linkage type and solubility	Specific activity ( $\mu\text{mol} \cdot \mu\text{mol}^{-1} \text{protein} \cdot \text{min}^{-1}$ )
Avicel (PH101)	$\beta$ -(1 $\rightarrow$ 4)-D-Glc, crystalline, insoluble	9.0 $\pm$ 0.7
Solka-Floc <sup>®</sup>	$\beta$ -(1 $\rightarrow$ 4)-D-Glc, crystalline, insoluble	5.3 $\pm$ 0.4
Regenerated cellulose (RAC)	$\beta$ -(1 $\rightarrow$ 4)-D-Glc, amorphous, insoluble	57.7 $\pm$ 4.0
Hydroxyethyl cellulose (HEC)	$\beta$ -(1 $\rightarrow$ 4)-D-Glc, soluble	875.0 $\pm$ 75.0
Carboxymethyl cellulose (CMC)	$\beta$ -(1 $\rightarrow$ 4)-D-Glc, partially substituted with carboxy methyl groups, soluble	1816.0 $\pm$ 149.0
$\beta$ -Glucan from barley grain	$\beta$ -(1 $\rightarrow$ 4)/(1 $\rightarrow$ 3)-D-Glc, soluble	2337.0 $\pm$ 164.0
Laminarin	$\beta$ -(1 $\rightarrow$ 3)-D-Glc backbone, mainly with $\beta$ -(1 $\rightarrow$ 6)-D-glucosyl substitutions, soluble	ND
Curdlan	$\beta$ -(1 $\rightarrow$ 3)-D-Glc, soluble	ND
Xyloglucan	$\beta$ -(1 $\rightarrow$ 4)-D-Glc backbone, mainly with $\beta$ -(1 $\rightarrow$ 6)-D-Xyl substitutions, soluble	14.1 $\pm$ 0.17
Glucomannan from Konjac	$\beta$ -(1 $\rightarrow$ 4)-D-Man/D-Glc, soluble	230.5 $\pm$ 21.4
Galactomannan from Locust bean gum	$\beta$ -(1 $\rightarrow$ 4)-D-mannan, single D-galactosyl units attached to C-6 of some of the D-mannosyl residues, soluble	2.8 $\pm$ 0.7
Wheat arabinoxylan	$\beta$ -(1 $\rightarrow$ 4)-D-Xyl backbone, partially substituted at O-2 and/or O-3 positions with L-Araf	ND
Colloidal chitin	$\beta$ -(1 $\rightarrow$ 4)-N-acetylglucosamine, amorphous, insoluble	ND

**Table 1.** Specific activity of the purified p4818Cel5\_2A enzyme towards various polysaccharides. Enzyme activities were measured with 1.0% substrates in 100 mM MES buffer pH 6.0 under 50 °C for 10 min. The amount of the purified p4818Cel5\_2A enzyme added for the activity assays on CMC,  $\beta$ -glucan from barley and HEC was 1.5  $\mu\text{g}$  protein, 30  $\mu\text{g}$  protein for xyloglucan, and 135  $\mu\text{g}$  protein for all other substrates. The incubation time was 30 min for Avicel and Solk-Floc<sup>®</sup>, 60 min for laminarin and chitin, and 10 min for other substrates. Details of the activity assay were described in the Material and Methods section. All values were expressed as means  $\pm$  SE, n = 3. ND: no detectable activity.

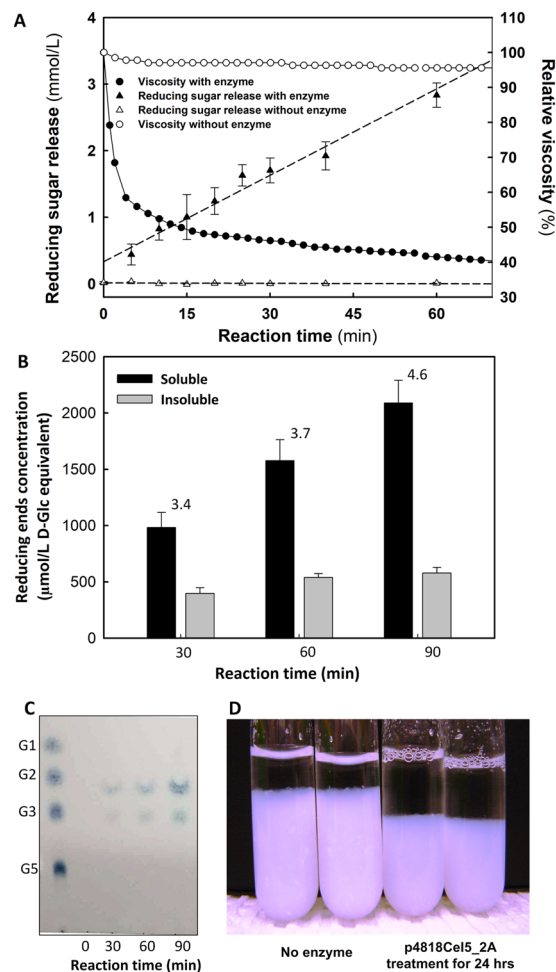
Polymeric substrate	$K_{m(\text{app})}$ (mg/ml)	$V_{\text{max}}$ ( $\mu\text{mol} \cdot \mu\text{mol}^{-1} \text{protein} \cdot \text{min}^{-1}$ )	$V_{\text{max}}/K_{m(\text{app})}$ ( $\mu\text{mol} \cdot \mu\text{mol}^{-1} \text{protein} \cdot \text{min}^{-1}) / (\text{mg/ml})$
Carboxymethyl cellulose (CMC)	10.9 $\pm$ 1.1	3540.0 $\pm$ 130.0	324.8
$\beta$ -Glucan from barley	0.94 $\pm$ 0.15	2664.0 $\pm$ 177.0	2834.0
Glucomannan from Konjac	4.4 $\pm$ 1.5	1353.0 $\pm$ 392.0	307.5
Synthetic substrate	$K_m$ (mM)	$K_{\text{cat}}$ ( $\text{min}^{-1}$ )	$K_{\text{cat}}/K_m$ ( $\text{min}^{-1} \cdot \text{mM}^{-1}$ )
pNP-Cellobioside	1.2 $\pm$ 0.1	97.8 $\pm$ 3.5	80.8
pNP- $\beta$ -D-Glucopyranoside	ND	ND	ND

**Table 2.** Kinetics of p4818Cel5\_2A towards soluble substrates. Activity assays were performed by the varying concentrations of corresponding substrates in 100 mM MES buffer pH 6.0 with an incubation of 10 min under 50 °C. The amount of the purified p4818Cel5\_2A enzyme added for the activity assays on CMC,  $\beta$ -glucan from barley was 1.5  $\mu\text{g}$  protein, and that for glucomannan from Konjac, pNP (4-Nitrophenol) - Cellobioside and pNP- $\beta$ -D-Glucopyranoside was 135  $\mu\text{g}$  protein respectively. Kinetic parameter estimates  $\pm$  SE (n = 3). ND: no detectable activity.

These tested fiber components are widely present in various animal diets and rations, such as forages, roughages, corn, barley, rice, wheat, as well as legumes (e.g. as peas, beans and soybean). Therefore, the broad substrate specificity indicates the involvement of p4818Cel5\_2A in digesting a wide variety of plant polysaccharides in pig hindgut, reflecting again a diet-driven enzyme evolution in shaping substrate specificity. In addition, the broad substrate specificity of p4818Cel5\_2A will be beneficial for its application as the exogenous feed enzyme in improving feed nutritional value.

**p4818Cel5\_2A shows a potentially processive mode of action.** In order to understand the action mode of p4818Cel5\_2A, its activity towards CMC was further examined by monitoring viscosity change and reducing sugar release over the reaction time. The viscosity of CMC assay solution declined rapidly by  $\sim$ 50% in the first 10 min after adding enzyme (Fig. 3A), and the entire viscosity change fitted well with a mono-exponential decay pattern ( $R^2 = 0.844$ ). In contrast, a linear increase for total reducing sugar release was observed over the reaction ( $R^2 = 0.963$ ) (Fig. 3A). Together, these results suggested that p4818Cel5\_2A cellulase acts as a  $\beta$ -(1  $\rightarrow$  4)-endoglucanase.

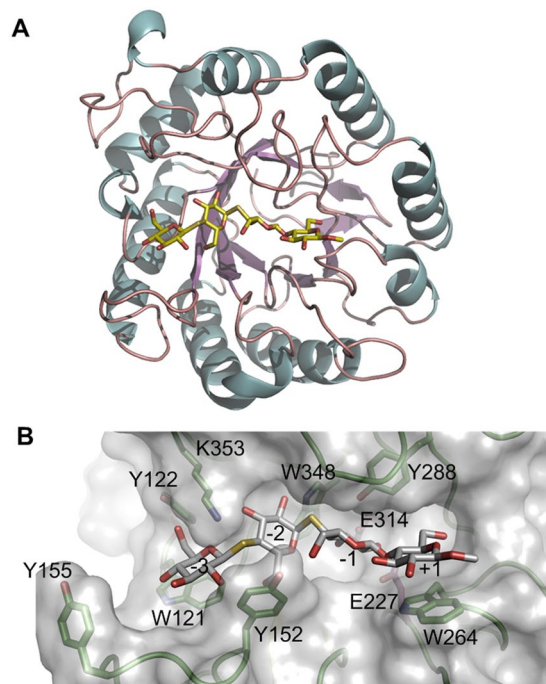
Processivity plays a critical role in cellulose enzymatic degradation, however, the assessment of cellulase processivity is challenging, as there is no straightforward approach to quantify it<sup>45</sup>. In this study, the processivity of p4818Cel5\_2A in hydrolyzing RAC was further examined by assessing the distribution of reducing sugar ends. Specifically, the ratio of reducing ends in the soluble fraction to that in the insoluble fraction (i.e. processive ratio) was increased from 3.4 to 4.6 as the incubation time was prolonged from 30 to 90 min (Fig. 3B), generating 77.3% and 82.1% soluble reducing sugar ends, respectively. This result stands in contrast to the actions of random cutting endoglucanases, which will produce more reducing sugar ends in the insoluble fraction than in the soluble fraction<sup>46</sup>. Further with thin layer chromatography (TLC), the predominant hydrolysis products



**Figure 3.** Enzyme p418Cel5\_2A acts as a processive  $\beta$ -1,4-endoglucanase. **(A)** The response of viscosity and reducing sugar release from carboxymethyl cellulose (CMC) solution hydrolyzed by p418Cel5\_2A. The purified enzyme (5.2  $\mu$ g) was added to 3% CMC solution in 100 mM MES buffer (pH 6.0) to initiate the reaction in a final volume of 16 ml, and the viscosity and reducing sugar concentration ( $\mu$ M) in the reaction solution was monitored at 5 min intervals for 1 hour. **(B)** The processivity of p418Cel5\_2A towards regenerated amorphous cellulose (RAC, 1.0%, wt/vol). The processivity is presented as the ratio of the reducing ends in soluble fraction to that in insoluble fraction, and labelled at top of the bars. **(C)** Thin layer chromatography (TLC) analysis of hydrolysis products from RAC by p418Cel5\_2A. G1, D-glucose; G2, cellobiose; G3, cellotriose; and G5, cellopentose. **(D)** Solubilization of RAC (1.0%, W/V) by p418Cel5\_2A for 24.0 h under 25 °C. Values were expressed as means  $\pm$  SE,  $n = 3$ .

of p418Cel5\_2A from RAC were identified as cellobiose and cellotriose with a very minor amount of glucose, irrespective of reaction time of 30, 60 or 90 min (Fig. 3C). In contrast, a wide range of cello-oligosaccharides are expected for the RAC hydrolysis by a typical random cutting endoglucanases. In total, 16.5% RAC (by dry matter) was solubilized by p418Cel5\_2A after 24 h digestion at 25 °C (Fig. 3D). Notably, cellobiose and cellotriose were also found to be the main hydrolysis products from RAC by several other processive GH5 endoglucanases<sup>39,41</sup>. Furthermore, p418Cel5\_2A also demonstrated activity towards *p*-nitrophenol (*p*NP)-cellobioside, a typical synthetic substrate for  $\beta$ -(1  $\rightarrow$  4)-D-glucan cellobiohydrolase activity<sup>47</sup>, with a  $K_m$  of  $1.2 \pm 0.1$  mM and  $k_{cat}$  of  $97.8 \pm 3.5$  min<sup>-1</sup> (Table 2). In contrast, p418Cel5\_2A did not show any detectable hydrolytic activity towards *p*NP- $\beta$ -D-glucopyranoside, which is a typical synthetic substrate for measuring  $\beta$ -glucosidase activities and is only one glucose unit shorter than *p*NP-cellobioside. Moreover, *p*NP- $\beta$ -D-glucopyranoside (10.0 mM) did not show any inhibitory effect to p418Cel5\_2A activity on *p*NP-cellobioside (1.0 mM), indicating a poor binding of *p*NP- $\beta$ -D-glucopyranoside to p418Cel5\_2A enzyme (Table 2). Together, these results suggested that the subsite -2 of p418Cel5\_2A (following the nomenclature defined by Davies and colleagues<sup>48</sup>) played an important role in the binding *p*NP-cellobioside to the enzyme (also see Fig. 4B). In summary, our results indicate that p418Cel5\_2A is a processive  $\beta$ -(1  $\rightarrow$  4)-endoglucanase for hydrolyzing cellulose into cellobiose and cellotriose as the primary end products.

In general, there are two types of processive cellulases, exocellulases and processive endoglucanases<sup>49</sup>. To date, most of reported processive endoglucanases are from GH families of 5, 9 and 48 whose members often exhibit multi-modular architectures, typically with a catalytic domain and one or more carbohydrate-binding modules



**Figure 4.** A homology model of p4818Cel5\_2A. The 3-dimensional structure model for the cellulase p4818Cel5\_2A was generated from residue 91 to 384 with SWISS-MODEL online server by using the crystal structure of the homologous cellulase (pdb#1E5J) as a template. **(A)** TIM barrel fold of p4818Cel5\_2A with an estimated overall diameter size of 50 Å. **(B)** Predicted cellulose binding cleft and relevant aromatic amino acids involved in substrate binding. The residues of Glu227 and Glu314 were predicted as the catalytic acid/base. The cellotetraose analogue of methyl 4<sup>II</sup>,4<sup>III</sup>-dithio- $\alpha$ -cellobiosyl-(1  $\rightarrow$  4)- $\beta$ -cellobioside was modelled by superimposing 1E5J and is shown as sticks; select residues believed to be important in substrate binding and catalysis are shown as green sticks.

(CBMs)<sup>50</sup>. Within the GH5 family, processivity has been reported for several multi-modular GH5\_2 endoglucanases, including a cellulase from *Bacillus subtilis* BS-5 (BsuEG5C, GH5-CBM3)<sup>41</sup>, an endoglucanase from *Saccharophagus degradans* (SdeCel5H, GH5-CBM6)<sup>39</sup> and an endoglucanase from *Hahella chejuensis* (HchCel5, CBM6-GH5)<sup>40</sup>, respectively (Fig. 1B). Of note, processive endoglucanases were also identified in the other sub-families of the GH5 family (e.g. GH5\_5 and GH5\_25), including a GH5\_5 endoglucanase from *Volvariella volvacea* (VvoEG1, CBM1-GH5)<sup>42</sup> and a GH5\_25 cellulase from *Thermotoga maritime* (TmaCel5A)<sup>51</sup> (Fig. 1B). In contrast, p4818Cel5\_2A is a single catalytic domain with a potential processive mode of action that hydrolyzes cellulosic substrates without any additional binding domains. To date, only a few of mono-modular cellulases have been reported to be processive, including the aforementioned TmaCel5A and a GH5\_2 endoglucanase from *Cytophaga hutchinsonii* (CHU-2103)<sup>52</sup> (Fig. 1B). To the best of our knowledge, p4818Cel5\_2A also represents the first microbial processive endoglucanase discovered from the monogastric animal gut microbiome, which is able to hydrolyze cellulose mainly to cellobiose and cellotriose. More interestingly, the classical scheme for cellulose microbial degradation involves the synergistic action of a minimum of three enzymes: randomly cutting  $\beta$ -(1  $\rightarrow$  4)-endoglucanase,  $\beta$ -(1  $\rightarrow$  4)-exoglucanase and  $\beta$ -glucosidase. With cellobiose and cellotriose as the main hydrolysis products, the processive  $\beta$ -(1  $\rightarrow$  4)-endoglucanase of p4818Cel5\_2A could potentially form a novel two-enzyme pathway for degrading cellulose into fermentable monosaccharide through further coupling with  $\beta$ -glucosidase or  $\beta$ -glycoside phosphorylase in pigs' gut.

**Structure model rationalized the monomodular catalytic features of p4818Cel5\_2A.** A structural model of p4818Cel5\_2A was built using the crystal structure of a homologous GH5\_2 cellulase (amino acid sequence identity: 47.0%) from *Salipaludibacillus agaradhaerens* (previously called *Bacillus agaradhaerens*) (BagCel5A, PDB ID#:1E5J) as template with a RMSD value of 0.354 Å<sup>53</sup>. This model displayed a classic triosephosphate isomerase (TIM) barrel fold with dimensions of 5.5 nm  $\times$  6.2 nm  $\times$  7.5 nm (Fig. 4A). Notably, the structure of BagCel5A showed a maximal diameter of 5.5 nm, minimum of  $\sim$ 4.5 nm. Erickson observed that most single modular proteins would fold into one globular domain and polypeptides larger than 50 kDa typically form two or more domains<sup>54</sup>. Thus, the monomodular cellulases with low molecular weight likely form various roughly spherical shapes with differing radius, while bi-modular, larger molecular weight cellulases tend to show ellipsoid or other irregular shapes<sup>55,56</sup>. Enzyme penetration into biomass is one of the limiting factors in biomass degradation<sup>57</sup>, where the pore size in substrates plays an important role. For example, the pore size in living plant cell wall was reported for a diameter ranging from 3.5 to 5.2 nm<sup>58</sup>, and the pretreated cellulosic substrates from lignocellulosic biomass materials have larger pore diameters, ranging between 5.1–11.0 nm<sup>59</sup>. The p4818Cel5\_2A is modeled to have an overall globular shape (Fig. 4A) with an estimated diameter of 4.6 nm<sup>79</sup>. Therefore, the



small size of the monomodular p4818Cel5\_2A enzyme relative to multi-modular enzymes will be beneficial for its ability to penetrate into lignocellulosic biomass, and could serve as an ideal industrial biocatalyst candidate for further biological engineering.

By comparing our model with the template structure (BagCel5A) in complex with a substrate analog<sup>53</sup>, a narrow and extended cellulose binding cleft was identified at the C-terminal side of the p4818Cel5\_2A TIM barrel (Fig. 4A). The presence of several aromatic residues, including the highly conserved W121, W348, W264 and Y288 along with Y152 and Y122, leads us to propose that this binding cleft contains at least four cellulose binding subsites (Figs 4B, S5). In particular, W264 and W121 are likely involved in hydrophobic stacking interaction with the glucose residues at subsites +1 and -3 respectively, while W348, Y152 and Y122 appear to contact glucose ring at subsite -2 or -3 through edge-on-edge interactions. Moreover, five out of the eight aforementioned amino acid residues (W121, Y122, Y152, Y158, Y290) lining substrate binding cleft are contributed by loop regions, indicating a potential flexibility of binding cleft conformation. This flexibility likely conferred p4818Cel5\_2A a broad substrate profile for cleaving  $\beta$ -(1  $\rightarrow$  4)-D-glucosyl bonds in a wide variety of plant polysaccharides (Tables 2, S4). In addition, the region of the binding cleft flanking the subsites from -3 to +1 is relatively open, which should allow the enzyme to accommodate xylosyl and galactosyl substitutions in the substrates xyloglucan and galactomannan, respectively. Interestingly, the side chains of Y152 and Y155 from loop region (residues 152 to 156) and K353 would further extend both sides of the substrate binding cleft in p4818Cel5\_2A at binding subsites of -3 and -2 (Fig. 4B). It would likely enhance the attachment of p4818Cel5\_2A to cellulose chain and allow enzyme to perform multiple rounds of catalysis before dissociating, in turn resulting in the observed processivity (Fig. 3A,B). Similarly, a well-characterized processive GH9 endoglucanase, TjCel9A, from *Thermobifida fusca* has been currently shown to processively move on a polysaccharide by utilizing six sugar binding subsites, which is independent from the associated CBM<sup>60</sup>. Notably, residues of Y152 and Y155 in p4818Cel5\_2A are not conserved across GH5 cellulases, and are often replaced by glycine, serine and leucine (Fig. S3, Table S4). A future study with structural characterization and site-specific mutagenesis will shed more light on the molecular mechanism of the catalytic features of p4818Cel5\_2A enzyme.

## Conclusions

Through previous dietary crystalline cellulose induction and metagenomic screening, a novel gut bacterial cellulase gene, p4818Cel5\_2A, identified from porcine gut microbiome, was further recombinantly expressed in the current study. This p4818Cel5\_2A cellulase was resistant to trypsin digestion. Our data further showed that p4818Cel5\_2A enzyme is broadly active towards a wide variety of plant polysaccharides, including cellulosic substrates of avicel and solka-Floc<sup>®</sup>, and the hemicelluloses of  $\beta$ -(1,4)/(1,3)-glucan, xyloglucan, glucomannan and galactomannan. Moreover, p4818Cel5\_2A acted as a processive  $\beta$ -1,4-endoglucanase for hydrolyzing cellulose into cellobiose and celotriose as the primary end products. It is monomodular with a modest total molecular weight, allowing it to form a compact domain. In summary, our study has revealed a novel processive  $\beta$ -1,4-endoglucanase with multi-substrate specificity derived from a porcine gut symbiont that enables facile cellulose digestion, but is also a promising biocatalyst candidate with the potential in multi-industrial biomass bioprocessing applications. The porcine gut microbiome has clear potential as a unique genomic resource for the discovery of biomass degradation enzymes.

## Materials and Methods

**Chemicals.** QIAquick PCR purification kit and QIAamp DNA stool mini kit were all purchased from QIAGEN (Toronto, ON, Canada). AZO-CM-cellulose (AZO-CMC), curdlan, xyloglucan and GH5 family cellulase from *Bacillus amyloliquefaciens* (BamGH5A) were from Megazyme (Wicklow, Ireland). Carboxymethyl cellulose (CMC, low viscosity), hydroethyl cellulose (HEC), Avicel (PH101), laminarin, xylan from birch wood, locust bean gum (LBG), pNP (*p*-nitrophenol)- $\beta$ -D-cellobioside, pNP- $\beta$ -D-glucopyranoside, IPTG, porcine trypsin and bovine chymotrypsin were all purchased from Sigma-Aldrich (Oakville, ON, Canada). pCR4Blunt-TOPO vector kit and Pfx Ultra DNA polymerase were from Invitrogen (Burlington, ON, Canada). Restriction enzymes and T4 DNA polymerase and ligase were from Fermentas (Burlington, ON, Canada). All other chemicals were of analytical grade and were obtained from either Sigma-Aldrich or Fisher Scientific (Nepean, ON, Canada). The regenerated amorphous cellulose (RAC) was prepared from the avicel (PH101) as previously described by Zhang *et al.*<sup>61</sup>

**Animal feeding experiment and digesta collection.** The crystalline cellulose pig feeding experiment for the induction of gut microbial cellulase genes had been conducted as described previously<sup>23</sup>. The induction diet including 10% crystalline cellulose of Solka-Floc<sup>®</sup> as the sole fiber source was adapted from Rideout *et al.* with modifications<sup>21</sup>. The details of this diet formulation have not yet been reported, and are now supplemented here for the first time (Table S2).

**Construction of metagenomic expression plasmid library.** Metagenomic DNA extraction, expression library construction and functional screening were performed as described previously<sup>23</sup>. The DNA insert size determination and the activity-based functional screening were further illustrated in the Supplemental Materials (Fig. S1). The inserts in positive clones were sequenced by primer walking and assembled in the Molecular Super Center at the University of Guelph. The nucleotide sequence of insert in positive clone of p4818 was deposited in GenBank with an accession number of MH373350.

**Sequence and phylogenetic analyses.** The possible open reading frames (ORFs) in the insert sequence were identified with the ORF finder at National Center for Biotechnology Information (NCBI; [www.ncbi.nlm.nih.gov/projects/gorf/](http://www.ncbi.nlm.nih.gov/projects/gorf/)). Modularity and signal peptides of the enzymes were predicted by Simple Modular Architecture Research Tool (SMART; <http://smart.embl-heidelberg.de>). The characterized GH5 enzyme

sequences (563, as of Mar. 3<sup>rd</sup>, 2019) were retrieved from the CAZy database (<http://www.cazy.org/>). The sequences were aligned using MUSCLE in Geneious 8.0.5 (Biomatters Ltd, New Zealand), and the sequences with fewer than 100 amino acids or that lack catalytic residues (i.e., equivalent to E314 and E227 in p4818Cel5\_2A) were removed. The catalytic domains of the remaining 556 curated GH5 sequences were extracted and then realigned using Geneious 8.0.5 prior to constructing a neighborhood-joining tree using Geneious Tree Builder. The bootstrap analysis was chosen for resampling with a replicate number of 100.

**Protein expression and purification.** The gene encoding the mature form of p4818Cel5\_2A (with the signal peptide sequence of 1–23 amino acids removed) was amplified by PCR using the primers with the sequences of GGTCACATATGTTAACTCTGGGGATTATAATTG (forward) and CTCCAAGCTTCTAGTGCTGGTTCAGTATGTC (Reverse) (introduced NdeI and HindIII restriction sites are underlined). The amplified fragment was purified, digested and ligated into the vector pET28a, and fused in frame with an N-terminal His-tag for generating pET28a-p4818Cel5\_2A using standard protocols<sup>62</sup>. The construct was verified by DNA sequencing in the Guelph Molecular Super Center.

Recombinant *E. coli* BL21 ( $\lambda$ DE3) cells harboring the overexpression construct of pET28a-p4818Cel5\_2A were propagated in 1 L of Luria-Bertani medium supplemented with 50  $\mu$ g/mL kanamycin at 37 °C until optical density at 600 nm ( $OD_{600nm}$ ) reached 0.8. The culture was then induced by adding IPTG to a final concentration of 0.5 mM and followed by shaking at 200 rpm for 18 h at 15 °C. The cells were harvested by centrifugation at  $5,000 \times g$  for 10 min. The cell pellet was resuspended in a lysis buffer containing 300 mM NaCl, 10 mM imidazole and 50 mM sodium HEPES at pH 7.0 and disrupted by a French Press three times at an operating pressure of 12000 psi. The cell debris was removed by centrifugation at  $17,500 \times g$  for 10 min. Unless stated otherwise, the buffer used for affinity chromatography is 50 mM HEPES containing 300 mM NaCl at pH 7.0. The supernatant was filtered through a 0.45- $\mu$ m filter and mixed with the Ni-NTA resin for 40 min with the gentle stirring under 4 °C. The mixture was poured into a column and the resin was washed with 300 ml of buffer containing 20 mM imidazole. The protein product p4818Cel5\_2A was eluted with a buffer containing 250 mM imidazole. The fractions with enzyme activity were pooled and dialyzed overnight with the storage buffer of 50 mM HEPES containing 5% (v/v) glycerol, pH 7.0, and then concentrated with Pierce protein concentrator (10 K; MWCO; Fisher Scientific; Ottawa, ON, Canada) to a final concentration of 3.5 mg/ml. The purified enzyme was aliquoted and stored at –80 °C.

**Enzyme activity assay.** All enzyme activity assays were performed in triplicate. The hydrolysis activity on various polysaccharides was examined by measuring reducing sugars released from substrates with the dinitro-salicylic acid (DNS) method<sup>63</sup>. Standard enzyme assay solutions contained 5.0 mM DTT and 1.0% (wt/vol) (10.0 mg/ml) of polysaccharide in 0.50 ml of 100 mM MES buffer (pH 6.0). The reaction was initiated by the addition of an amount of enzyme determined to release products in a linear relation to time when incubated at 50 °C for up to 10 min. The reaction was terminated by adding 1.5 ml of DNS solution and followed by incubation of 10 min under 100 °C to develop the color. The absorbance at 540 nm was measured using Pharmacia Ultrospec 2000 UV/VIS Spectrophotometer. One enzyme activity unit (U) was defined as the amount of enzyme that releases 1.0  $\mu$ mole of D-(+)-glucose equivalent from the tested substrate per minute under activity assay conditions. D-(+)-glucose was used to generate a standard curve.

The p4818Cel5\_2A activity on 4-nitrophenyl (pNP) glycosides was determined by measuring pNP release<sup>61</sup>. The typical reaction mixture contained 1.0 mM of the substrate in 0.5 ml of 100 mM MES buffer (pH 6.0). After the reaction was initiated by the addition of enzyme, the reaction mixture was incubated at 50 °C for 10 min. The reaction was terminated by adding 1.0 ml of 4.0% (wt/v)  $Na_2CO_3$  to the mixture. The amount of pNP formed was measured by the  $A_{405nm}$  (extinction coefficient  $\epsilon[405\text{ nm}] = 17,600\text{ M}^{-1}\text{ cm}^{-1}$ ), and pNP was used to generate a standard curve. As indicated above, the enzyme dosage added to each assay mixture was optimized to measure initial reaction rates.

The kinetic assays were carried out under standard conditions except that the substrate concentrations were varied from at least 0.1  $K_m$  to 5  $K_m$ . The collected data were fitted by the nonlinear regression analysis to the Michaelis-Menten equation using Sigmaplot 10.0.

The hydrolysis pattern of the enzyme was also investigated by monitoring specific viscosity decrease in a 3.0% (wt/vol) (30.0 mg/ml) CMC solution using a controlled strain rheometer (MCR 301, Anton Paar GmbH, Ostfildern, Germany) equipped with a Peltier temperature controller and a parallel plate geometry (PP50,  $d = 50$  mm, Anton Paar) was used. The purified enzyme of 19.3  $\mu$ g was added to 30 mg/ml CMC solution in 100 mM MES buffer at pH 6.0 to trigger the reaction in a final volume of 16 ml, and the viscosity and reducing sugar concentration of the reaction solution were monitored at 2.0 and 5.0 min intervals, respectively, for 1.0 h.

**Optimum enzyme reaction conditions, enzyme stability, and effects of divalent ions.** The effect of pH on enzyme activity was determined over a pH range from 4.0 to 10.0 with increments of 0.5 pH-unit using a constant ionic strength buffer containing 100 mM Tris, 50 mM acetic acid and 50 mM MES. The same buffers were used to examine the pH effect to the stability of enzyme activity, and the enzyme was incubated at different pH levels at 4 °C for 24 h prior to the activity assay. The temperature effect on enzyme activity and stability was assessed at different temperatures ranging from 15 to 70 °C in 50 mM HEPES buffer, pH 7.0. The incubation time under different temperatures was 60 min. CMC 1.0% (wt/vol) was used as the substrate for the standard enzyme assays described above.

**Resistance of the enzyme activity to trypsin and chymotrypsin digestion.** The p4818Cel5\_2A or BamCel5 (Megazyme, E-CELBA) in the final concentration of 50  $\mu$ g/ml was incubated with either 5000 U/ml trypsin [the Sigma/Aldrich BAEE enzyme activity unit using Na-benzoyl-L-arginine ethyl ester (BAEE) as a typical assay substrate] or 200 U/ml chymotrypsin [the Sigma/Aldrich BTEE enzyme activity unit using

N-benzoyl-L-tyrosine ethyl ester (BTEE) as a typical assay substrate], respectively, at 37 °C in 100 mM MES buffer (pH 6.0). The residual cellulase activity was monitored at different incubation periods by the standard activity assay with CMC as a substrate.

**Cellulase processivity and hydrolysis product analysis.** Cellulase processivity was presented as the ratio of soluble reducing ends to insoluble reducing sugar ends produced in assay reaction when using the regenerated amorphous cellulose (RAC) as a substrate<sup>45</sup>. Briefly, 1% (wt/vol, 10.0 mg/ml) RAC in 0.4 mL 100 mM MES buffer pH 6.0 (8.0 mg FP/ml) was incubated with 0.43 nmol p4818cel5\_2A (18.8 µg) under 35 °C for the incubation time of 30, 60 and 90 min, respectively. The reaction was terminated by adding 0.05 ml 2.0 M NaOH and heat treatment at 70 °C for 10 min to ensure inactivation of the enzyme, and then neutralized using 0.05 ml of 2.0 M HCl. After centrifugation at 6900 × g, 10 min, the supernatant was quickly transferred and its reducing sugars were measured by DNS method with D-glucose as a standard as aforementioned<sup>63</sup>. The insoluble pellet was washed with 6.0 M guanidine HCl in 0.6 mL twice to remove any bound protein first and further with 100 mM MES buffer (pH 6.0) twice. MES buffer 0.5 mL was added back to washed filter paper disc. The remaining reducing ends in this insoluble fraction were then measured with DNS reagent. The hydrolysis products in the supernatant were analyzed with thin-layer chromatography (TLC) on Silica Gel 60 aluminum plates (Merck, Darmstadt, Germany), which was developed in ethyl acetate/water/methanol (8:3:4, v/v/v). The plate was sprayed with a diphenylamine-aniline stain<sup>64</sup>, which was made by dissolving 2 g of diphenylamine and 2 ml of aniline in 100 ml of acetone followed by the addition of 10 ml of concentrated phosphoric acid. After staining, plates were developed in an oven at 100 °C for up to 1 h.

**Protein assay.** Protein concentrations were determined by the Bradford assay using bovine serum albumin as a standard. SDS-PAGE was performed and stained with Coomassie Blue according to established procedures<sup>65</sup>.

**3-D Structure homological modeling.** The 3-dimensional model of residue 91 to 384 of cellulase p4818Cel5\_2A was generated by the SWISS-MODEL online server<sup>66</sup> using the crystal structure of a homologous cellulase (PDB ID:1E5J) as a template<sup>53</sup>. Structure model was analyzed and visualized using PyMOL 2.1 ([www.pymol.org](http://www.pymol.org)).

### Data Availability

All data generated or analyzed during this study are included in this published article and its Supplementary Information Files.

### References

- Chundawat, S. P. S., Beckham, G. T., Himmel, M. E. & Dale, B. E. Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* **2**, 121–45 (2011).
- Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat. Rev. Microbiol.* **6**, 121–31 (2008).
- Sekirov, I., Russell, S. L., Antunes, L. C. M. & Finlay, B. B. Gut microbiota in health and disease. *Physiol. Rev.* **90**, 859–904 (2010).
- Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–141 (2014).
- Fan, M. Z. *Concepts of Animal Nutrition, the 2nd Custom Edition*. (Pearson, 2015).
- Agyekum, A. K. & Nyachoti, C. M. Nutritional and Metabolic Consequences of Feeding High-Fiber Diets to Swine: A Review. *Engineering* **3**, 716–725 (2017).
- Bhat, M. K. Cellulases and related enzymes in biotechnology. *Biotechnol. Adv.* **18**, 355–383 (2000).
- Fan, M. Z. Swine nutrition and environment. In *Sustainable Swine Nutrition* 365–411 (John Wiley & Son Inc., 2013).
- Daniel, R. The metagenomics of soil. *Nat. Rev. Microbiol.* **3**, 470–478 (2005).
- Warnecke, F. *et al.* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**, 560–565 (2007).
- Hess, M., Sczyrba, A., Egan, R., Kim, T.-W. & Chokhawala, H. Metagenomic discovery of biomass-degrading genes and genomes from bow rumen. *Science* **331**, 463–467 (2011).
- Beloqui, A. *et al.* Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. *Appl. Environ. Microbiol.* **76**, 5934–5946 (2010).
- Ferrer, M. *et al.* Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ. Microbiol.* **7**, 1996–2010 (2005).
- Tasse, L. *et al.* Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* **20**, 1605–1612 (2010).
- Nyyssönen, M. *et al.* Coupled high-throughput functional screening and next generation sequencing for identification of plant polymer decomposing enzymes in metagenomic libraries. *Front. Microbiol.* **4**, 1–14 (2013).
- Lewin, A. *et al.* Novel archaeal thermostable cellulases from an oil reservoir metagenome. *AMB Express* **7** (2017).
- Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3**, 289–306 (2012).
- Morrison, M., Pope, P. B., Denman, S. E. & McSweeney, C. S. Plant biomass degradation by gut microbiomes: more of the same or something new? *Curr. Opin. Biotechnol.* **20**, 358–363 (2009).
- White, B. A., Lamed, R., Bayer, E. A. & Flint, H. J. Biomass utilization by gut microbiomes. *Annu. Rev. Microbiol.* **68**, 279–296 (2014).
- Kidder, D. & Manners, M. Passage of Food. In *Digestion in the pig* (Scientific, Kinston Press, 1978).
- Rideout, T. C., Liu, Q., Wood, P. & Fan, M. Z. Nutrient utilisation and intestinal fermentation are differentially affected by the consumption of resistant starch varieties and conventional fibres in pigs. *Br. J. Nutr.* **99**, 984–992 (2008).
- Navarro, D. M. D. L., Bruininx, E. M. A. M., Jong, L. D. & Stein, H. H. The contribution of digestible and metabolizable energy from high-fiber dietary ingredients is not affected by inclusion rate in mixed diets fed to growing pigs. *J. Anim. Sci.* **96**, 1860–1869 (2018).
- Wang, W. *et al.* The porcine gut microbial metagenomic library for mining novel cellulases established from growing pigs fed cellulose-supplemented high-fat diets. *J. Anim. Sci.* **90**, 400–402 (2012).
- Yang, H. *et al.* Uncovering the composition of microbial community structure and metagenomics among three gut locations in pigs with distinct fatness. *Sci. Rep.* **6**, 27427 (2016).
- Gerlt, J. A. *et al.* The enzyme function initiative. *Biochemistry* **50**, 9950–9962 (2011).
- Koeck, D. E., Pechtl, A., Zverlov, V. V. & Schwarz, W. H. Genomics of cellulolytic bacteria. *Curr. Opin. Biotechnol.* **29**, 171–183 (2014).

27. Lynd, L. R., Weimer, P. J., Zyl, W. H. V. & Pretorius, I. S. Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiol. Mol. Biol. Rev.* **66**, 506–577 (2002).
28. Stapleton, M. *et al.* Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches. *Curr. Opin. Microbiol.* **18**, 1–7 (2014).
29. Hengge-Aronis, R. Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol. Mol. Biol. Rev.* **66**, 373–95 (2002).
30. Landick, R., Wade, J. T. & Grainger, D. C. ScienceDirect H-NS and RNA polymerase: a love – hate relationship? *Curr. Opin. Microbiol.* **24**, 53–59 (2015).
31. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H. & Henriksat, B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* **12** (2012).
32. Park, B. H., Karpinets, T. V., Syed, M. H., Leuze, M. R. & Uberbacher, E. C. CAZymes Analysis Toolkit (cat): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* **20**, 1574–1584 (2010).
33. Romaniec, M. P., Huskisson, N., Barker, P. & Demain, A. L. Purification and properties of the *Clostridium thermocellum* bglB gene product expressed in *Escherichia coli*. *Enzyme Microb. Technol.* **15**, 393–400 (1993).
34. Banwell, J. G., Leonard, P. J. & Lobo, R. M. Measurement of trypsin and chymotrypsin activity in stools to detect chronic pancreatic disease. *Gut* **6**, 143–5 (1965).
35. Tanabe, T., Morinaga, K., Fukamizo, T. & Mitsutomi, M. Novel Chitosanase from *Streptomyces griseus* HUT 6037 with Transglycosylation Activity. *Biosci. Biotechnol. Biochem.* **67**, 354–364 (2003).
36. Zhang, Y. H. P. & Lynd, L. R. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnol. Bioeng.* **88**, 797–824 (2004).
37. Qi, M., Jun, H. S. & Forsberg, C. W. Characterization and synergistic interactions of *Fibrobacter succinogenes* glycoside hydrolases. *Appl. Environ. Microbiol.* **73**, 6098–6105 (2007).
38. Wilson, D. B. Studies of *Thermobifida fusca* plant cell wall degrading enzymes. *Chem. Rec.* **4**, 72–82 (2004).
39. Watson, B. J., Zhang, H., Longmire, A. G., Young, H. M. & Hutcheson, S. W. Processive endoglucanases mediate degradation of cellulose by *Saccharophagus degradans*. *J. Bacteriol.* **191**, 5697–5705 (2009).
40. Ghatge, S. S. *et al.* Characterization of modular bifunctional processive endoglucanase Cel5 from *Hahella chejuensis* KCTC 2396. *Appl. Microbiol. Biotechnol.* **98**, 4421–4435 (2014).
41. Wu, B. *et al.* Processivity and enzymatic mechanism of a multifunctional family 5 endoglucanase from *Bacillus subtilis* BS-5 with potential applications in the saccharification of cellulosic substrates. *Biotechnol. Biofuels* **11**, 1–15 (2018).
42. Zheng, F. & Ding, S. Processivity and enzymatic mode of a glycoside hydrolase family 5 endoglucanase from *Volvariella volvacea*. *Appl. Environ. Microbiol.* **79**, 989–996 (2013).
43. Zarafeta, D. *et al.* Discovery and characterization of a thermostable and highly halotolerant GH5 cellulase from an icelandic hot spring isolate. *PLoS One* **11**, 1–18 (2016).
44. Alvarez, T. M. *et al.* Structure and function of a novel cellulase 5 from sugarcane soil metagenome. *PLoS One* **8**, e83635 (2013).
45. Horn, S. J., Sørli, M., Vårum, K. M., Våljamäe, P. & Eijsink, V. G. H. Measuring processivity. *Methods Enzymol.* **510**, 69–95 (2012).
46. Zverlov, V. V., Schantz, N. & Schwarz, W. H. A major new component in the cellulosome of *Clostridium thermocellum* is a processive endo- $\beta$ -1,4-glucanase producing cellotetraose. *FEMS Microbiol. Lett.* **249**, 353–358 (2005).
47. McCleary, B. V., McKie, V. & Draga, A. Measurement of endo-1,4- $\beta$ -glucanase. *Methods Enzymol.* **510**, 1–17 (2012).
48. Davies, G. J., Wilson, K. S. & Henriksat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* **321**(Pt 2), 557–559 (1997).
49. Wilson, D. B. & Kostylev, M. *Cellulase processivity*. **908**, 93–99 (2012).
50. Horn, S. J. *et al.* Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides. *Proc. Natl. Acad. Sci. USA* **103**, 18089–94 (2006).
51. Basit, A. & Akhtar, M. W. Truncation of the processive Cel5A of *Thermotoga maritima* results in soluble expression and several fold increase in activity. *Biotechnol. Bioeng.* **115**, 1675–1684 (2018).
52. Zhang, C. *et al.* Characterization of a multi-function processive endoglucanase CHU-2103 from *Cytophaga hutchinsonii*. *Appl. Microbiol. Biotechnol.* **98**, 6679–6687 (2014).
53. Fort, S. *et al.* Mixed-linkage cellooligosaccharides: A new class of glycoside hydrolase inhibitors. *ChemBioChem* **2**, 319–325 (2001).
54. Erickson, H. P. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol. Proced. Online* **11**, 32–51 (2009).
55. Cowling, E. B. & Kirk, T. K. Properties of cellulose and lignocellulosic materials as substrates for enzymatic conversion processes. *Biotechnol Bioeng Symp.*, 95–123 (1976).
56. Abuja, P. M., Hayn, M. & Chen, H. The structure of endoglucanase I (*Trichoderma resei*) in solution. *Prog. Colloid Polym. Sci.* **181**, 8010 (1993).
57. Ding, S. How does plant cell wall nanoscale architecture correlate with enzymatic digestibility. *Science (80-)*. **338**, 1055–1061 (2012).
58. Lin, K. W., Ladisch, M. R., Voloch, M., Patterson, J. A. & Noller, C. H. Effect of pretreatments and fermentation on pore size in cellulosic materials. *Biotechnol. Bioeng.* **27**, 1427–1433 (1985).
59. Grethlein, H. E. The effect of pore size distribution on the rate of enzymatic hydrolysis of cellulosic substrates. *Nat. Biotechnol.* **155**–160 (1985).
60. Sørli, M., Hamre, A. G., Kaupang, A., Payne, C. M. & Va, P. Thermodynamic Signatures of Substrate Binding for Three *Thermobifida fusca* Cellulases with Different Modes of Action; <https://doi.org/10.1021/acs.biochem.9b00014> (2019).
61. Zhang, Y. H. P., Hong, J. & Ye, X. Cellulase assays. *Methods Mol. Biol.* **581**, 213–231 (2009).
62. Sambrook, J., Maniatis, T. & Fritsch, E. *Molecular cloning: A laboratory manual*. (Cold Spring Harbor Laboratory Press, 1989).
63. Miller, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**, 426–428 (1959).
64. Bailey, R. W. & Bourne, E. J. Colour reactions given by sugars and diphenylamine-aniline spray reagents on paper chromatograms. *J. Chromatogr. A* **4**, 206–213 (1960).
65. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
66. Biasini, M. *et al.* SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, 252–258 (2014).
67. Solovyev, V. & Salamov, A. Automatic annotation of microbial genomes and metagenomic sequences. in *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies* (ed. Li, R. W.) 61–78 (Nova Science Publishers, 2011).

## Acknowledgements

We thank Dr. Julang Li, Dr. Stephen Seah, Dr. Emma Allen-Vercoe and Dr. Milena Corredig from University of Guelph and Dr. Emma Master from University of Toronto for their generous technical supports and insightful discussions. We also thank Dr. Bernard Henriksat at Centre National de la Recherche Scientifique, Marseille, France for his kind advice on the subfamily classification of this GH5 sequence. This study was supported by projects from the Natural Science and Engineering Research Council (NSERC) of Canada, the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) New Directions & Alternative Renewable Fuels Research

Program; the University of Guelph-OMAFRA Partnership Research Program, the University of Guelph Gryphon's LAAIR (Leading to Accelerated Adoption of Innovative Research) program, Agriculture, Agri-Food Canada (AAFC) Swine Innovation Porc (SIP) Swine Cluster Program, and the Metagen Enzyme Corporation (to M.Z.F.).

### Author Contributions

M.Z.F. obtained research fund, as well as conceived and coordinated the study; W.W. and T.A. performed the study; W.W., J.S.L., M.S.K. and M.Z.F. wrote the paper. All authors reviewed the results and approved the final version of the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-50050-1>.

**Competing Interests:** Ming Z. Fan (M.Z.F.) is the co-founder and board chairman of the Metagen Enzyme Corporation, and has an active grant-in-aid research funding at the University of Guelph from the Metagen Enzyme Corporation. Weijun Wang (W.W.) and Tania Archbold (T.A.) are board members of the Metagen Enzyme Corporation. The other authors declare that no competing interests exist.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019