JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Determinants of Regioselective Hydroxylation in the Fungal Polysaccharide Monooxygenases

Van V. Vu,[†] William T. Beeson,[‡] Christopher M. Phillips,[§] Jamie H. D. Cate,[‡,§,∥,⊥] and Michael A. Marletta*,[†]

[†]Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, United States

[‡]Department of Chemistry, [§]Department of Molecular and Cell Biology, and [∥]California Institute for Quantitative Biosciences, University of California at Berkeley, Berkeley, California 94720, United States

[⊥]Division of Physical Biosciences, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States

Supporting Information

ABSTRACT: The ubiquitous fungal polysaccharide monooxygenases (PMOs) (also known as GH61 proteins, LPMOs, and AA9 proteins) are structurally related but have significant variation in sequence. A heterologous expression method in *Neurospora crassa* was developed as a step toward connecting regioselectivity of the chemistry to PMO phylogeny. Activity assays, as well as sequence and phylogenetic analyses, showed that the majority of fungal PMOs fall into three major groups with distinctive active site surface features. PMO1s and PMO2s hydroxylate glycosidic positions C1 and C4, respectively. PMO3s hydroxylate both C1 and C4. A subgroup of PMO3s (PMO3*) hydroxylate C1. Mutagenesis studies showed that an extra subdomain of about 12 amino acids contribute to C4 oxidation in the PMO3 family.

Polysaccharide monooxygenases (PMOs) are ubiquitous in filamentous fungi.[1−4] They have been recently found to play a central role in oxidative degradation of cellulose[1,5−11] and to have potential in the conversion of plant biomass to commodity chemicals such as biofuels. Crystal structures of several PMOs reveal an absolutely conserved copper-bound active site coordinated by two His residues on a flat protein surface.[1,5,12−14] PMOs are proposed to bind and then cleave glycosidic bonds of cellulose, generating new chain ends on the cellulose surface that are subsequently processed by other cellulases (Figure 1A). It is generally accepted that PMOs activate O_2 at their copper center to hydroxylate cellulose resulting in the elimination of the glycosidic bond. In this reaction, four electrons are required per O_2 molecule utilized, with two provided by the substrate and the other two delivered from either chemical reductants[1,6,7,15] or cellobiose dehydrogenases (CDHs)[6,16] (Figure 1B). Features other than the copper active site on the flat surface vary between PMOs,[1,5,12−14] which likely influence the orientation of the copper center toward different oxidation targets on cellulose, resulting in regioselective oxidation as suggested by computational studies.[12,13]

Product characterization is one of the fundamental aspects of the PMO reaction that needs to be resolved. Aldonic acids, 4-ketoaldoses, and 6-hexodialdoses have been reported as the corresponding products of oxidation at C1, C4, and C6, respectively. While aldonic acids are well characterized, characterization of the other two types of products is more challenging due to the lack of synthetic standards. Based on mass spectrometry analysis, 6-hexodialdoses were reported as the products of *Thermoascus aurantiacus* GH61A (*Ta*GH61A)[5] and *Podospora anserina* GH61B (*Pa*GH61B).[11] However, the formation of 6-hexodialdoses, which have the same molecular weight as that of the corresponding 4-ketoaldoses, has not been rigorously proven. In contrast, comprehensive chemical, high performance anion exchange chromatographic, and mass spectrometric analyses showed that *Neurospora crassa* PMO1 and PMO2 produced aldonic acids and 4-ketoaldoses, respectively, while 6-hexodialdoses were not detected (Figure 1B).[7]

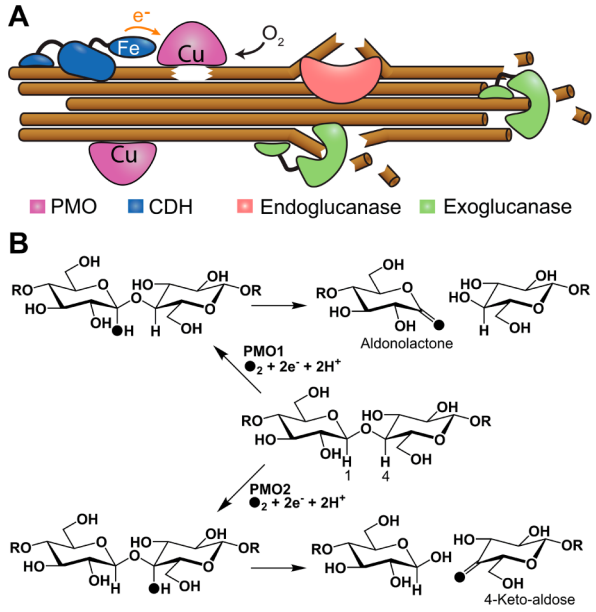

Figure 1. Oxidative cellulose degradation by PMOs. (A) A working model of cellulose degradation. (B) Glycosidic bond cleavage via C1 and C4 hydroxylation by *Nc*PMO1 and *Nc*PMO2.

Received: September 10, 2013
Published: December 18, 2013

   dx.doi.org/10.1021/ja409384b | *J. Am. Chem. Soc.* 2014, 136, 562−565

To obtain deeper insight into the PMO family, we aimed to classify them into groups based on sequence analysis and regioselectivity. A phylogenetic tree of cellulose-upregulated PMOs from *Neurospora crassa* and related enzymes was built using PhyML3.0. The unrooted tree in Figure 2 shows three major clades. NCU08760 and Pc_GH61D have been shown to form aldonic acid products[6,7,13] and belong to the same clade (orange leaves). NCU01050 oxidizes C4 and belongs to the same clade as NCU02240 and NCU02916 (blue leaves). The third clade contains Ta_GH61A (turquoise). These three branches are designated as PMO1, PMO2, and PMO3, respectively.

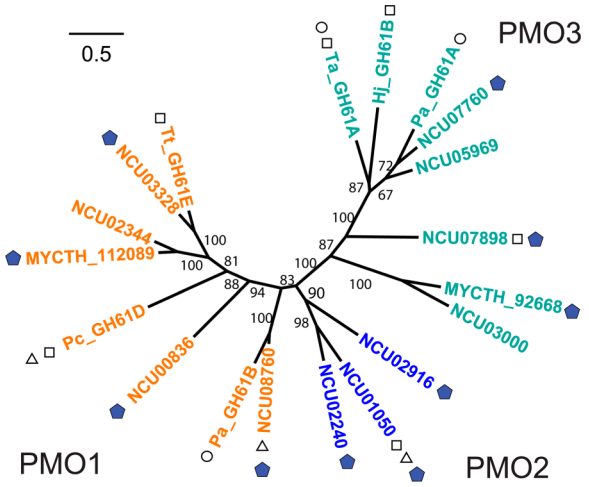

Figure 2. Unrooted phylogenetic tree of cellulose-upregulated *Neurospora crassa* PMOs and related enzymes. (□) Crystallized. (△) Products verified. (○) Products not verified. (Blue pentagon) Studied here, among which NCU08760, NCU02240, NCU01050, and NCU07898 purified natively and the rest expressed recombinantly. NCU08760 was also expressed recombinantly. Support values (%) are indicated next to the branches.

In order to determine the regioselectivity of each PMO group, representative PMOs of each group were expressed. Native PMOs undergo several co- and post-translational modifications in their native host, including signal peptide cleavage to generate the N-terminal histidine residue, N-methylation of this residue, and O- and N-linked glycosylation. Therefore, heterologous expression of PMOs in bacteria has not been successful. Although expression of functional PMOs is possible in *Pichia pastoris*, the methylation of the N-terminal histidine was not reported.[9,11,17] The lack of a direct comparison of the activity of recombinant PMOs expressed in *P. pastoris* and that of natively purified enzymes limits the interpretation of *P. pastoris*-expressed proteins. Alternatively, expression of PMOs in filamentous fungi, such as *Aspergillus oryzae*,[1] can potentially carry out all the post-translational modifications. We developed a recombinant expression method in *N. crassa*. A glyceraldehyde-3-phosphate dehydrogenase promoter was inserted upstream of the target PMO gene (Scheme S1). The construct was then introduced to the *N. crassa* chromosome via replacement of the *csr-1* gene as described previously.[18] Positive *N. crassa* transformants express the target PMOs as the only secreted cellulase upon growth on glucose-supplemented minimal media. Native *N. crassa* cellulases are not expressed under these conditions.

Recombinant and native PMOs were purified as described in the supplemental methods.[6] ESI mass spectra of recombinant NCU08760 (rNCU08760) exhibit similar mass envelopes to that of natively purified NCU08760 (Figure S2), indicating that the recombinant enzyme is correctly processed. LC/MS analysis of trypsin-digested recombinant PMOs confirmed methylation of N-terminal histidine (data not shown). Furthermore, the activity of rNCU08760 is the same as that of natively purified NCU08760 in assays with phosphoric acid swollen cellulose (PASC) (Figure S3).

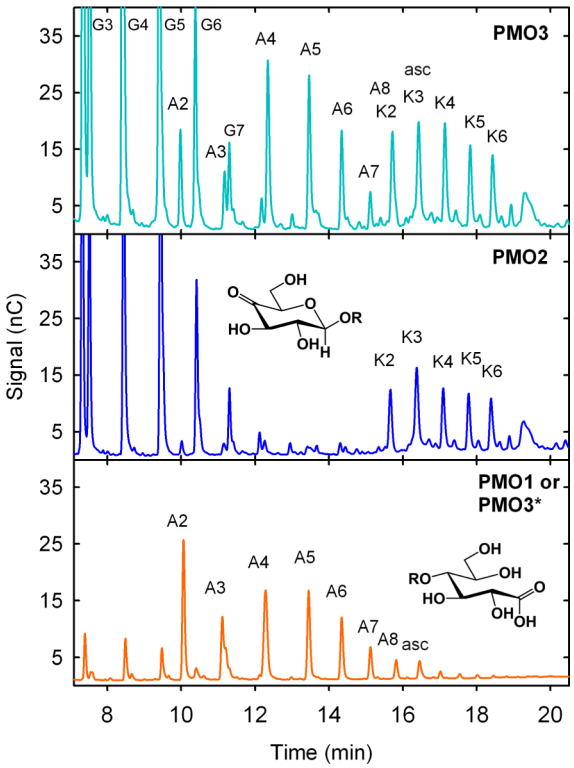

Figure 3. Representative HPAEC traces of PMO (~5 μM) activity assays on PASC (5 mg/mL) at pH 5.0 and 40 °C with ascorbic acid (2 mM) as the exogenous reductant. G3−G7 = cellodextrins with degrees of polymerization of 3−7; A2−8 = aldonic acids; K2−6 = 4-ketoaldoses; asc = ascorbic acid.

Activity assays with PASC were carried out for all recombinant PMOs and the four natively purified *N. crassa* PMOs. Soluble reaction products were analyzed with high performance anion exchange chromatography (HPAEC). Figure 3 shows representative HPAEC traces for each type of PMO in Figure 2. A complete set of chromatograms is provided in the Supporting Information (Figure S4). For PMO1s, a series of aldonic acid peaks of increasing chain lengths are observed. These aldonic acid peaks are not reliably detected in the product chromatographic traces of PMO2s. Instead, a series of new peaks appear at later retention times, which are attributed to 4-ketoaldoses as previously shown[7] (Figures 3 and S4). In the PMO3 branch, only aldonic acids were observed for MYCTH_92668, while both aldonic acids and 4-ketoaldoses are detected for NCU07898 and NCU07760.

The identities of the reaction products were further confirmed using a previously established procedure:[7] (1) Reduction of 4-ketoaldoses with sodium borohydride followed by hydrolysis with trifluoroacetic acid (TFA) results in the

dx.doi.org/10.1021/ja409384b | J. Am. Chem. Soc. 2014, 136, 562−565

formation of a mixture of galactose, glucose, and sorbitol, while galactose will not be observed for the reduction of hexodialdoses and aldonic acids; (2) Oxidation of hexodialdoses with hypoiodite followed by TFA hydrolysis yields both gluconic and glucuronic acids while glucuronic acid will not be observed in the case of 4-ketoaldoses and aldonic acids. This procedure was used to analyze the soluble products of five recombinant PMOs and four natively purified *N. crassa* PMOs reported previously (Figure S5). A mixture of galactose, glucose, and sorbitol was observed for all three PMO2s and the two PMO3s NCU07760 and NCU07898. Galactose was not detected for any PMO1s and the PMO3MYCTH_92668. Glucuronic acid was not detected for any of the PMOs tested.

Taken together with phylogenetic analysis, the activity assays and product identification results indicate that PMOs can be divided into three groups based on chemical regioselectivity and the evolutionary relationship. PMO1s and PMO2s are specific for C1 and C4 oxidation, respectively, consistent with previous observation.[7] However, as shown here some PMO3s hydroxylate both the C1 and C4 positions. The implications of these results are described below.

The branching pattern of the small set of PMOs in Figure 2 is consistent with that of a larger number of PMOs. Harris and colleagues reported a cladogram of about 300 PMOs in 2010 that can essentially be divided into three major branches.[1] We found at least 500 sequences from NCBI and UniProtKB protein databases with conserved PMO motifs, including the *N*-terminal His residue, the [Hx$_n$GP] motif for the second His ligand of the copper active site, and the [QxYxxC] motif for the active site tyrosine residue. The phylogenetic tree built from these sequences (Figure 4) fills in the branching pattern as that of the small tree (Figure 2). The two unlabeled branches have yet to be characterized. These branches account for approximately 10% of the PMOs shown in this tree.

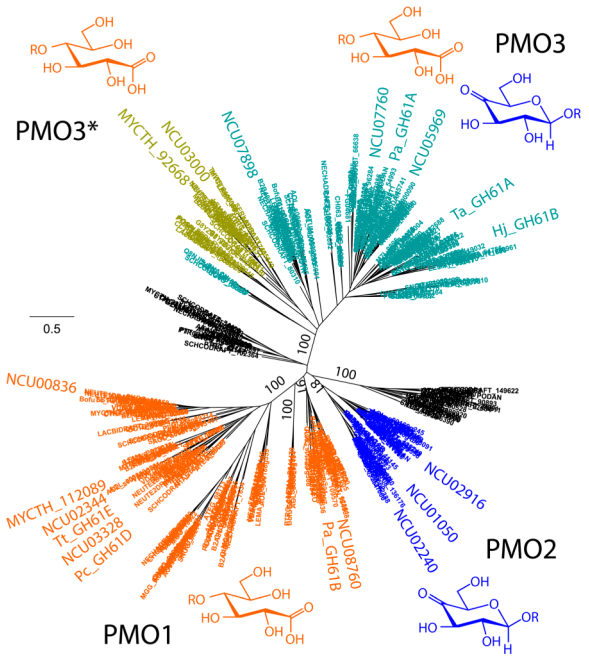

Figure 4. Unrooted phylogenetic tree of 497 PMOs. Respective oxidation products are indicated next to PMO groups. Black: unclassified due to the lack of characterized representatives. For an extended view, see Figure S6.

Based on our results, it is likely that the majority of fungal PMOs oxidize the C1 and/or C4 position in cellulose. C6 oxidation was not observed for any PMOs studied here. While cleavage of the glycosidic bond is facilitated by C1 and C4 oxidation as shown in Figure 1B,[6,7] oxidation at the C6 position does not have this outcome. The phylogenetic analysis and sequence alignment (*vide infra*) suggests that Ta_GH61A and Pa_GH61A are closely related to NCU07760 (Figure 2) and are predicted to hydroxylate both C1 and C4 as found for NCU07760. Likewise, Pa_GH61B is closely related to NCU08760 and is predicted to hydroxylate C1 as observed for NCU08760.

Previous structural and computational studies suggested that specific active site surface features may contribute to the regioselectivity of PMOs.[12,13] To identify group-specific active site surface features we analyzed multiple sequence and structural alignments (Figures 5 and S7). A distinctive feature of PMO3s is an insert of approximately 12 amino acids starting around residue 20. Crystal structures of NCU07898,[12] Ta_GH61A,[1] and Hj_GH61B[14] show that this inserted sequence forms a helix that lies parallel to the flat active site surface (Figure 5).[5,12] The NCU07760 mutant lacking this insert (residues 21−32) produces aldonic acids predominantly (Figure S8), indicating that this insert is important for C4 selectivity in PMO3s.

A subgroup of PMO3s, designated PMO3*, contains MYCTH_92668 that only produces aldonic acids. The extra insert in the PMO3* subgroup lacks several conserved amino acids found in PMO3s, which might result in the loss of the short helix found in PMO3s. The insert in PMO3* does not contain the conserved Y24 in PMO3s and other aromatic residues (Figures 5 and S7). Tyrosine is commonly found in recognition motifs of carbohydrate binding proteins.[19] The aromatic side chain of Y24 aligns along the active site surface, suggesting that Y24 may be the key residue for C4 oxidation by PMO3s. However, mutation of Y24 in the PMO3 NCU07760 to glycine does not alter the regioselectivity (Figure S8). Thus, other residues or motifs within the extra insert of PMO3s appear to govern C4 oxidation in this group.

Another distinctive feature of PMO3s and PMO2s is a conserved surface S residue that is immediately before the second copper active site histidine ligand (Figures 5 and S7). This surface S residue is replaced by a T residue in PMO3*s and is not present in PMO1s. The crystal structures of NCU01050, NCU07898, and Ta_GH61A show that the S hydroxyl group is positioned on the active site surface, which could contribute to interactions with the cellulose substrate (Figure 5). PMO2s also have a highly conserved lysine residue immediately following the second histidine ligand. In addition, there is an extra unique sequence of 9−14 amino acids a few residues before the second His ligand in PMO2s. The crystal structure of NCU01050 shows that this sequence also forms a short helix on the active site flat surface,[12] which may directly interact with the cellulose substrate, although it does not contain aromatic residues (Figures 5 and S7).

In summary, through the characterization of various PMOs heterologously expressed in *Neurospora crassa*, we have been able to connect regioselectivity to phylogeny. Based on our analysis, the majority of fungal PMOs found in protein databases can be divided into major groups that selectively oxidize either or both glycosidic C1 and C4. Through structural and sequence alignments, we have identified features on the active site surface that likely contribute to fine-tuning the

dx.doi.org/10.1021/ja409384b | J. Am. Chem. Soc. 2014, 136, 562−565

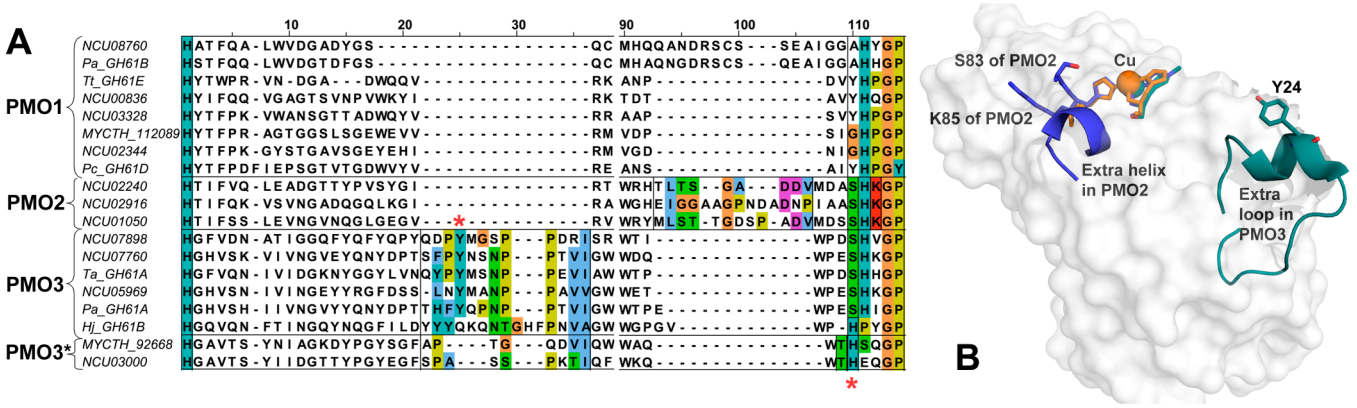

Figure 5. (A) Multiple sequence alignment (MSA) of 19 PMOs listed in Figure 2. An MSA of 497 PMOs in Figure 4 is provided in Figure S6. (B) Comparison of distinctive features on the flat active site surface of PMO1 (orange) (3EII), PMO2 (4EIR) (blue), and PMO3 (2YET) (turquoise).

orientation of the copper active site on the cellulose surface for regioselective hydroxylation. We have shown that an extra subdomain of about 12 amino acids plays a key role in regioselectivity at C4. Further mutagenesis studies involving the modifications of group-specific surface features and structural analysis are underway.

■ ASSOCIATED CONTENT

Supporting Information

Experimental methods, additional data, and multiple sequence alignment and phylogenetic tree in rooted circular format for 497 PMOs. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

marletta@scripps.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Elise Span and other Marletta group members for experimental support and helpful discussions. This work was funded by grants from the Energy Biosciences Institute to J.H.D.C. and M.A.M. and support from The Scripps Research Institute to M.A.M.

■ REFERENCES

(1) Harris, P. V.; Welner, D.; McFarland, K. C.; Re, E.; Navarro Poulsen, J.-C.; Brown, K.; et al. *Biochemistry* **2010**, *49*, 3305−3316.

(2) Tian, C.; Beeson, W. T.; Iavarone, A. T.; Sun, J.; Marletta, M. A.; Cate, J. H. D.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 22157−22162.

(3) Yakovlev, I.; Vaaje-Kolstad, G.; Hietala, A.; Stefanczyk, E.; Solheim, H.; Fossdal, C. G. *Appl. Microbiol. Biotechnol.* **2012**, *95*, 979−990.

(4) Berka, R. M.; Grigoriev, I. V.; Otillar, R.; Salamov, A.; Grimwood, J.; Reid, I.; et al. *Nat. Biotechnol.* **2011**, *29*, 922−927.

(5) Quinlan, R. J.; Sweeney, M. D.; Lo Leggio, L.; Otten, H.; Poulsen, J.-C. N.; Johansen, K. S.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 15079−15084.

(6) Phillips, C. M.; Beeson, W. T.; Cate, J. H. D.; Marletta, M. A. *ACS Chem. Biol.* **2011**, *6*, 1399−1406.

(7) Beeson, W. T.; Phillips, C. M.; Cate, J. H. D.; Marletta, M. A. *J. Am. Chem. Soc.* **2012**, *134*, 890−892.

(8) Horn, S.; Vaaje-Kolstad, G.; Westereng, B.; Eijsink, V. G. H. *Biotechnol. Biofuels* **2012**, *5*, 45.

(9) Westereng, B.; Ishida, T.; Vaaje-Kolstad, G.; Wu, M.; Eijsink, V. G. H.; Igarashi, K.; et al. *PLoS ONE* **2012**, *6*, e27807.

(10) Langston, J. A.; Shaghasi, T.; Abbate, E.; Xu, F.; Vlasenko, E.; Sweeney, M. D. *Appl. Environ. Microbiol.* **2011**, *77*, 7007−7015.

(11) Bey, M.; Zhou, S.; Poidevin, L.; Henrissat, B.; Coutinho, P. M.; Berrin, J.-G.; et al. *Appl. Environ. Microbiol.* **2013**, *79*, 488−496.

(12) Li, X.; Beeson, W. T.; Phillips, C. M.; Marletta; Michael, A.; Cate, J. H. D. *Structure* **2012**, *20*, 1051−1061.

(13) Wu, M.; Beckham, G. T.; Larsson, A. M.; Ishida, T.; Kim, S.; Payne, C. M.; et al. *J. Biol. Chem.* **2013**, *288*, 12828−12839.

(14) Karkehabadi, S.; Hansson, H.; Kim, S.; Piens, K.; Mitchinson, C.; Sandgren, M. *J. Mol. Biol.* **2008**, *383*, 144−154.

(15) Cannella, D.; Hsieh, C.-w.; Felby, C.; Jorgensen, H. *Biotechnol. Biofuels* **2012**, *5*, 26.

(16) Sygmund, C.; Kracher, D.; Scheiblbrandner, S.; Zahma, K.; Felice, A. K. G.; Harreither, W.; et al. *Appl. Environ. Microbiol.* **2012**, *78*, 6161−6171.

(17) Kittl, R.; Kracher, D.; Burgstaller, D.; Haltrich, D.; Ludwig, R. *Biotechnol. Biofuels* **2012**, *5*, 79−92.

(18) Bardiya, N.; Shiu, P. K. T. *Fungal. Genet. Biol.* **2007**, *44*, 307−314.

(19) Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem. J.* **2004**, *382*, 769−781.