Enzymatic breakdown of lignocellulosic biomass: the role of glycosyl hydrolases and lytic polysaccharide monooxygenases

Uchenna Regina Ezeilo, Iffah Izzati Zakaria, Fahrul Huyop & Roswanira Abdul Wahab

To cite this article: Uchenna Regina Ezeilo, Iffah Izzati Zakaria, Fahrul Huyop & Roswanira Abdul Wahab (2017) Enzymatic breakdown of lignocellulosic biomass: the role of glycosyl hydrolases and lytic polysaccharide monooxygenases, Biotechnology & Biotechnological Equipment, 31:4, 647-662, DOI: 10.1080/13102818.2017.1330124

To link to this article: https://doi.org/10.1080/13102818.2017.1330124

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 16 May 2017.

Submit your article to this journal

Article views: 2636

View Crossmark data

Citing articles: 2 View citing articles
Review: Agriculture and Environmental Biotechnology

Enzymatic breakdown of lignocellulosic biomass: the role of glycosyl hydrolases and lytic polysaccharide monoxygenases

Uchenna Regina Ezeilo, Iffah Izzati Zakaria, Fahrul Huyop and Roswanira Abdul Wahab

Department of Biotechnology and Medical Engineering, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia; Department of Chemistry/Biochemistry/Molecular Biology, Federal University Ndufu-Alike Ikwo, Ebonyi, Nigeria; Natural Products and Drug Discovery Center, Malaysian Institute of Pharmaceuticals and Nutraceuticals, National Institutes of Biotechnology Malaysia, Ministry of Science, Technology and Innovation, Pinang, Malaysia; Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

ABSTRACT
Lignocellulose constitutes a major component of discarded wastes from various industries viz. agriculture, forestry and municipal waste treatment. The potential use of lignocellulose from such types of biomass can be maximized by enzymatic degradation using glycoside hydrolases (GHs) and oxidative enzymes to produce renewable fuels. Nonetheless, besides the slow rate of degradation and low yields, lignocellulose is also physicochemically recalcitrant and costly to process, further limiting its mass utilization. Therefore, bioprospecting for micro-organisms producing efficient lytic polysaccharide monooxygenases (LPMOs) to overcome these drawbacks may prove beneficial. The use of GHs and LPMOs can potentially help to circumvent some limitations in the conversion of lignocellulosic biomass into fermentable sugars. LPMOs are classified as family GH61 or family 33 carbohydrate-binding module (CBM33), whose unusual surface-exposed active site is bound to a copper (II) ion. To date, there are more than 20 known genes encoding cellulase-active LPMOs in bacteria and fungi, with diverse biological activities. Only by thorough comprehension of the diversity, enzymology and role of primary GHs, i.e. cellulases and their oxidative machinery can the degradation of lignocellulosic biomass be improved. This review provides insight into the diversity, structure and mechanisms, structural and functional aspects of the oxidative breakdown of cellulose by LPMOs of the cellulase-active GH family.

KEYWORDS
Biofuels; cellulase; glycosyl hydrolases; lignocellulosic degradation; lytic polysaccharide monooxygenase; LPMO

INTRODUCTION
Lignocellulose offers the largest inexpensive and renewable source of potentially degradable carbohydrate on earth [1,2]. Unfortunately, the full potential of such biomass is normally underutilized and mostly wasted in the form of pre- and post-harvest agricultural losses and wastes from the food-processing industry [3]. According to the review of literature, the available lignocellulosic feedstock yielded from agriculture and other sources amounts to an approximate 180 million tons per year [4]. The lignocellulosic biomass primarily consists of polysaccharide polymers, cellulose and hemicellulose, and the phenolic polymer lignin [5,6]. However, cellulose, being the most abundant constituent, is structurally entrapped by the other cell-wall components, which hinders its enzymatic breakdown [7].

The structural robustness of cellulose can be attributed to the way the biopolymer is organized in large crystals containing tens of thousands of glucose molecules. The complex hydrogen-bonding (H-bonding) network found within and between glucan chains combined with the degree of polymerization involving thousands of glucose monomers further add to the limited accessibility of the glycolytic linkages to the hydrolytic enzymes [8]. Since the cellulose crystals are embedded (but not covalently linked to) a matrix of hemicellulose and lignin, this renders the access of biodegrading enzymes very difficult. Nevertheless, the recent discoveries of a diverse glycoside hydrolase family of enzymes and their oxidative auxiliaries promise huge potential applications of such enzymes for depolymerization of lignocellulosic biomass for bioconversion into value-added products.

In this review, we focus on the current standing of glycoside hydrolase families, their new class of copper-dependent lytic polysaccharide monooxygenases (LPMOs), highlighting their discovery, protein structure, predicted role and mechanism of catalysis. Their

CONTACT
Uchenna Regina Ezeilo ucmgbenka@gmail.com; Roswanira Abdul Wahab roswanira@kimia.fs.utm.my

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
potential applications in crop waste biomass degradation for biofuel production are also discussed.

**Enzymatic hydrolysis of cellulosic biomass**

Cellulose is a very stable molecule and it has been estimated that the uncatalysed hydrolytic degradation of such component would result in a half-life as long as 5 million years [9,10]. Thus, cellulose-degrading systems mediated by cellulases are necessary for industrial cellulose breakdown as well as to sustain the natural global carbon cycle. Cellulases are a class of enzymes produced mainly by fungi, bacteria and protozoans specifically for the cellulolysis (or hydrolysis) of biomass, i.e. cellulose [11]. As a matter of fact, the role of cellulases in the breakdown of cellulose has been investigated for both complex enzyme mixtures and individual components [12]. So far, the accepted model for enzymatic degradation of cellulose has been based on hydrolytic cellulase enzymes [13]. The revised model of Elwyn et al. [14] has shown that cellulose hydrolysis takes place in three concurrent steps: (1) physical and chemical changes in the yet unhydrolysed solid substrate; (2) primary hydrolysis in which soluble cello-oligomers are released from the solid cellulose to the hydrolysate and (3) secondary hydrolysis, in which the dissolved oligomers are hydrolysed to glucose [15].

Cellulases are members of the glycosidase hydrolase (GH) families of enzymes that catalyse the hydrolysis of β-1,4-glycosidic bonds of cellulose to glucose [13,16]. The cellulose-degrading enzymes are further divided into three major groups: endo-β-glucanases (EG), exoglucanases (celllobiohydrolases, CBH) and β-glucosidases (BGL), which belong to the EC 3.2.1.X class. They include endo-β-(1,4)-glucanases (EC 3.2.1.4), exo-β-(1,4)-glucanase (EC 3.2.1.91), exo-β-(1,4)-glucosidase (EC 3.2.1.74) and β-glucosidase (EC 3.2.1.21) [17–19] (Table 1). Figure 1 depicts the hydrolytic breakdown of cellulose by cellulases. The active sites of cellulases are found as a cleft or tunnel shape lined with aromatic residues whose role is to enhance the release of a glucan chain from cellulose [22,23]. It is hypothesized that the limiting step in the cellulase-catalysed breakdown of crystalline cellulose is the detachment of the glucan chain from the strong H-bonding network in cellulose into the active site groove [24].

Table 1. Cellulases, functions and their corresponding glycosyl hydrolase enzyme families.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>EC number</th>
<th>Functions</th>
<th>Glycosyl hydrolase (GH) Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Endo-β-(1,4)-glucanases (1,4): β-D-glucan-4-glucanohydrolase</td>
<td>EC 3.2.1.4</td>
<td>Randomly hydrolyses β-(1,4)-glycosidic bonds exposing reducing and non-reducing ends of the linear polymer of glucose units</td>
<td>GH5, GH9, GH12, GH44, GH45, GH48, GH51, GH74, GH124 [20]</td>
</tr>
<tr>
<td>2a. Exo-β-(1,4)-glucanases (1,4): β-D-glucan glucohydrolase</td>
<td>EC 3.2.1.74</td>
<td>Cleaves glucose successively from the non-reducing ends of the glucan to liberate glucose units, with preference for substrates of longer chain length and giving rise to inverted products</td>
<td>GH3, GH5, GH6, GH7, GH9, GH48 [20]</td>
</tr>
<tr>
<td>2b. Exo-β-(1,4)-glucanase (1,4): β-D-glucan cellobiohydrolase</td>
<td>EC 3.2.1.91</td>
<td>Acts at the reducing ends of β-(1,4)-glucans, produced by endoglucanases, cleaving cellobiose and cellooligosaccharides</td>
<td>GH5, GH6, GH7, GH9, GH48 [20]</td>
</tr>
<tr>
<td>3. β-glucosidase (β-D-glucoside glucohydrolase)</td>
<td>EC 3.2.1.21</td>
<td>Hydrolyses cellobiose and other very short chain β-D-oligosaccharides up to cellohexaose to form glucose; unlike exoglucosidases the rate of hydrolysis decreases markedly as the degree of polymerization of the substrate increases.</td>
<td>GH1, GH3, GH5, GH9, GH30, GH116 [20]</td>
</tr>
</tbody>
</table>

Note: Refs. [19] and [21].

Figure 1. Enzymatic cellulose degradation model.
**Glycosyl hydrolase enzyme diversity**

Glycosyl hydrolases (GH), also known as glycosidases, are a group of enzymes that hydrolyse glycosidic bonds between two or more sugars or a sugar and a non-sugar moiety within carbohydrates or oligosaccharides [25]. The enzymes within this family are widely distributed across prokaryotic, eukaryotic and archaea species [26–30] and have been reported to demonstrate interesting functional diversity as well as variation in copy number among organisms. To date, a total of 115 GH families have been identified based on their modes of action and amino-acid sequence. However, in recent times, due to the availability of more information on the protein structure and functions of such enzymes, it became clear that classification simply based on substrate-specificity was unsuitable. This is in light of the fact that similar protein folds may often exhibit several types of substrate specificities [30]. The recommendation of Cantarel et al. [28] on a classification method based on the effect of protein folding guided by the amino-acid sequence was found more suitable in assigning enzymes into their families and sub-families.

**Carbohydrate-binding module**

Generally, the use of GHs enzymes alone is inadequate for the breakdown of insoluble polysaccharides owing to difficulties in the enzymes’ access to the specific position of the substrate during catalysis. To overcome this problem, the GH catalytic modules are usually appended to one or more carbohydrate-binding modules (CBMs) capable of degrading insoluble polysaccharides. CBMs are the non-catalytic part of polysaccharide-degrading enzymes, such as cellulases and hemicellulases, that bind to the cell-wall polymers. In fact, the term CBM was suggested as a more inclusive term to designate all of the non-catalytic sugar-binding modules derived from GHs. Although many of these modules target components of the plant cell wall, several CBM families contain proteins that bind to insoluble storage polysaccharides, such as starch and glycogen. The role of CBMs has been known for enhancing the binding of the enzymes to the cellulose substrates [31], an aspect pertinent in efficient catalysis and degradation of unwanted biomass, such as agricultural-based biomass.

According to the review of literature, in recent years, the carbohydrate-active enzymes (CAZy) database groups the CBMs into 81 different families [29,32], based on amino-acid sequence similarity. Similar to the catalytic modules of GHs, 54 of the CBM families have been classified into Types A, B and C. CBMs of Type A, perhaps, are the most distinct among the CBMs, as this class only specifically binds to the surface of insoluble highly crystalline polysaccharides, such as cellulose and/or chitin [33]. The Type B CBMs function to identify internal glycan chains (endo-type) and interact only with single polysaccharide chains that bind to polysaccharides which constitute the substrates for the cognate catalytic module of the enzyme. Examples of enzymes appended to Type B CBMs are cellulases, xylanases and mannanases. In contrast, Type C CBMs link to the termini of glycans (exo-type). This unique class of CBMs has a ‘lectin-like’ feature that binds optimally to mono-, di- or tri-saccharides [34]. This is in agreement with their ability to recognize small sugars; hence their well-known ‘small-sugar-binding’ capacity, unlike those seen in Type B CPMs, which exhibit higher binding affinity towards longer oligosaccharide ligands [32,35].

**CBMs and glycoside hydrolase linkers**

Thorough comprehension of the structural biology by which CBMs bind to their target ligands may provide invaluable insights into the mechanisms of carbohydrate–protein recognition. Currently, there are several mechanisms that elucidate the ligand-binding between CBMs, enzymes as well as linkers. An earlier mechanism proposed by Creagh et al. [36] described that such ligand-binding involves entropically driven expulsion of water molecules from hydrophobic surfaces of cellulose and protein, typically shown by the binding between a CBM33 from the β-1,4-exoglucanase Cex of *Cellulomonas fimi* and insoluble bacterial microcrystalline cellulose. However, the molecular basis of the thermodynamically driven forces to explain the protein–carbohydrate-binding remains highly contentious.

Another mechanism, on the other hand, proposed that the cooperative-binding between the CBM and hydrolase was essential for carbohydrate recognition. An example can be seen in the modular enzyme Xyn10B from *Clostridium thermocellum* consisting of an N-terminal family 22 CBM (CBM22-1), a family 10 GH catalytic domain (GH10), another CBM22 module (CBM22-2), a dockerin sequence and a C-terminal family 1 carbohydrate esterase (CE1) catalytic domain. Residues from helix H4 of the GH10 module initiate the main contacts by binding into the minor groove of the CBM22-1 module [37]. The CBM22-1 orients in such a way that permits the substrate to bind loosely and be subsequently conveyed to the active site progressively [37], while another study reported that, by binding to a family 1 CBM, the thermostability of the GH10 xylanase from *Talaromyces cellulolytica* was improved [38]. It has been suggested that the family 10 CBM in the C-terminus of thermostable endo-β-1,4-xylanase (Xyl10A) plays two
major roles in the synergistic hydrolysis of lignocellulose by Xyl10A and cellulases. The binding of Xyl10A to family 10 CBM was seen to enhance lignocellulosic xylan hydrolysis via attachment to cellulose, as well as facilitating efficient removal of xylan obstacles that impede cellulase activity (because of the similar binding target of CBM1) [39]. Consequently, the combination of CBM-containing cellulases and xylanases in fungal systems could contribute to reduction of the enzyme loading in the hydrolysis of pretreated lignocellulose [38]. Likewise, CBMs have been found useful in the non-hydrolytic substrate disruption (amorphogenesis) [40] as well as in assisting the amorphogenesis of non-hydrolytic proteins [39]. Interestingly, CBMs can be multi-modulated with several lignocellulose-degrading catalytic domains by tethering to flexible glycosylated linkers [41]. Certain linkers have been shown to substantially improve binding as compared to only CBM. This results in effective binding of the substrate, an aspect that, presumably, improves the enzyme activity [42].

**Cellulosomes: multienzyme complexes**

Bioprospecting for effective cellulose degrading microorganisms led to the discovery of cellulosomes in the early 1980s. Cellulosomes are multienzyme complexes [43] produced mostly by anaerobic bacteria and fungi such as Clostridium and Ruminococcus spp., and Chytridomycetes [44], respectively. It is apparent that these proteins, or nanomachines [45], are important for efficient degradation of cellulose and hemicellullose [46]. The architecture of cellulosomes includes enzyme subunits such as endoglucanases, xylanases and cellobiohydrolases. These enzymes display varying substrate specificities and catalytic mechanisms coordinated by the scaffoldin protein.

Apart from the catalytic module, the dockerin module modulates enzyme interaction alongside the scaffoldin protein. Scaffoldins are multidomain and multifunctional proteins that boost interactions between catalytic proteins with the GH dockerin domains, consequently improving the affinity of the enzyme complex and its catalytic efficiency via CBMs [47] (Figure 2). The cohesion module sited on the scaffoldin binds to a dockerin module on each enzymatic subunit [48,49]. This architecture enables synergistic and well-coordinated enzymatic interactions, making cellulosomes about the most adequate biochemical system for the degradation of cellulose [20]. The past decade has seen artificial cellulosomes being constructed for the sole purpose of improving enzymatic functions for the saccharification reaction [50]. The construction was carried out via a chemical approach whereby a multicellulase conjugate was assembled on a double-stranded DNA scaffoldin. The resulting complex DNA–(endoglucanase)₃ conjugate exhibits a unique hydrolytic activity on crystalline cellulose (Avicel). The activity of the enzyme conjugate is dependent on the cellulase/DNA ratio of the DNA-based artificial cellulosome [51].

**Reaction steps in enzymatic lignocellulosic breakdown**

Enzymatic hydrolysis that converts lignocellulosic biomass to fermentable sugars involves complex steps. It has been extensively described that the enzymatic hydrolysis process is influenced by both the structural features of cellulose and the mode of enzyme action. Complete degradation of lignocellulosic biomass generally involves different sets of hydrolytic enzymes, such as cellulases, hemicellulases and other accessory enzymes [52]. To improve the lignocellulosic biomass...
breakdown for maximum biomass conversion, research efforts have been directed mainly on substrate-related factors, which include crystallinity, degree of polymerization, accessibility, preparation and properties of model substrates, and pretreated lignocellulosic materials [53].

Substrate-related factors

Structurally, the cellulose polymer is highly heterogeneous, which is why the hydrolysis of this biopolymer by cellulases incurs a highly complex process. Hence, establishing rational models based on mechanistic steps can be rather problematic due to the inherent complexities in both the substrate and enzyme [54,55]. The structure of cellulose consists of sugar rings in different chains, aligned on the same plane interacting with other layers of cellulose chains, forming the intra- and inter-chain hydrogen bond network [56] that gives cellulose its highly organized and stable, tightly packed structure, i.e. ‘crystalline’ [57]. Additionally, cellulose components include amorphous regions at varying sizes and accessibility as well as degradation rates. While the above-mentioned factors may differ from one carbohydrate source to another, the structurally relevant parameters, viz. chain length, crystallinity and number of accessible binding sites, can change with the progression of degradation [54,55].

It has been described that the efficiency of enzymatic hydrolysis of lignocellulose is affected by: (1) accessibility, (2) availability of surface area, (3) crystallinity, (4) degree of polymerization, (5) lignin and hemicellulose content, (6) changes in feature during degradation and, finally, (7) the pretreatment process [15]. Aside from adsorption, other factors such as diffusion, desorption and unproductive binding of different enzymes on different heterogeneous substrates should also be considered [58–60].

Enzyme-related factors

The exploitation of microbial cellulases for agricultural and industrial processes has been reported since the 1990s [21]. In general, cellulases are comprised of a mixture of several enzymes, three being most common in the hydrolysis process: (1) endoglucanase (EG, endo-1,4-D-glucanohydrolase, or EC 3.2.1.4.), which attacks regions of low crystallinity in the cellulose fibre, creating free chain-ends; (2) exoglucanase, or cellobiohydrolase (CBH, 1,4-b-D-glucan cellobiohydrolase, or EC 3.2.1.91.), which degrades the molecule further by removing cellobiose units from the free chain-ends and (3) β-glucosidase (EC 3.2.1.21), which hydrolyses cellobiose to produce glucose as the substrate for a myriad of manufacturing processes [61]. Studies that focus on biospecting for new micro-organisms have highlighted that both bacteria and fungi are good producers of cellulases, which includes Clostridium, Cellulomonas, Bacillus, Thermomonospora, Ruminococcus, Bacteriodes, Erwinia, Acetovibrio, Microbispora, Streptomyces, Cellulomonas fimii and Thermomonospora fusca, as well as the cellulolytic anaerobes such as C. thermocellum and Bacteroides cellulosolvens [62].

However, depending on the type of micro-organisms producing cellulases, the two most common cellulase systems are: (1) non-complexed cellulase system (usually associated with aerobic bacteria and fungi) and (2) complexed cellulase system (usually associated with anaerobic micro-organisms) [63]. The cellulases produced by the genus Trichoderma have received intensive attention due to the high levels of secreted cellulase; thus, the most fully investigated non-complexed cellulase system is the Trichoderma reesei model. Trichoderma reesei (teleomorph Hypocre a jecorina) is a saprobic fungus, known as an efficient producer of extracellular enzymes [64], which includes two cellobiohydrolases, at least seven endoglucanases and several glucosidases. The three hydrolytic processes to degrade cellulose carried out by T. reesei occur simultaneously and ultimately produce glucose as the final product [41]. On the other hand, the production of cellulosomes is usually associated with anaerobic bacteria (complexed cellulose system), which give some advantages: (1) synergism of the cellulases; (2) absence of unspecific adsorption [58].

Enzymatic lignocellulosic breakdown

Enzyme diffusion into (void of) solid substrates has been investigated and modelled, and the premise on which the model [59] was formed has been solved. The solved conditions include that (1) the enzymatic hydrolysis of lignocellulose depends on particle size tunable to a preferred size by pretreatment as well as (2) dramatic decrease in reaction rate after the initial burst. Thus, the influence of diffusion is downplayed.

The adsorption of cellulases is mediated by CBMs, type-specific for crystalline or amorphous regions only. Crystalline-specific CBMs are inclined to clustering on ridges, linear regions of glucan chains aligned parallel to one another. An example is the Trichoderma reesei cellobiohydrolase I (TrCBH) that binds preferentially to the hydrophobic (‘planar’) face of crystalline cellulose microfibrils [65]. The CBH1-binding sites are limited on crystalline cellulose due to the tight packing of chains as well as inaccessible chain-ends buried within the crystals. In contrast, amorphous-specific CBMs bind more uniformly across the surface, binding tightly to exposed glycans.
chains [65]. Likewise, endoglucanases bind specifically to more amorphous regions [54,55] of cellulose.

**GH families involved in lignocellulosic biomass degradation**

A diverse number of GH families contain the majority of enzymes that can catalyse LGC biomass degradation. Table 1 shows the cellulase enzymes and their respective GH families. Some of the largest groups are discussed below.

**GH family 3**

The GH family 3 (GH3) is one of the most abundant ones in the CAZy database, comprising over 6000 enzymes extensively distributed in plants, fungi and bacteria. The family exhibits various activities such as exo-acting β-D-glucosidases, α-L-arabinofuranosidases, β-D-xylopyranosidases and N-acetyl-β-D-glucosaminidases [66], all of which use a retaining glycosidase mechanism [20,67–69]. In addition to hydrolytic activities, some GH3 enzymes can catalyse glycosidic bond formation either via thermodynamically controlled reverse hydrolysis, or kinetically controlled transglycosylation [70,71]. In all, GH3 enzymes catalyse a range of functions, including cellulolic biomass degradation, plant and bacterial cell-wall remodelling, energy metabolism and pathogen defence [67,72]. These enzymes also have important roles in many other biological processes such as synthesis of functional glycosides from glycoside precursors [73] and cyanide-based biological defence mechanisms in plants [74].

With regards to substrate specificities, the GH3 family has extensive substrate-specificity with respect to monosaccharide residues, linkage position and chain length of the substrate. GH3 β-D-glucan glucohydrolases are also broadly specific exohydrolases that remove single glucosyl residues from the non-reducing ends of a range of β-D-glucans, (1,6)-β-D-glucosides and aryl β-D-glucosides. They include 1,3-β-D-glucans, 1,4-β-D-glucans, (1,3:1,4)-β-D-glucans and (1,6)-β-D-glucans, 4-nitrophenyl-β-D-glucoside, certain cyanogenic β-D-glucosides and some β-D-oligoxylglucosides [75]. In contrast to the broad substrate-specificity of the GH3 enzymes described above, there are exceptions such as the GH3 N-acetyl-β-D-glucosamine (GlcNAc) of *Cellulomonas fimi* Nag3 [76,77].

GH3 glycoside hydrolases act via a classical Koshland double-displacement mechanism, in which glycosyl residues are singly removed from the non-reducing ends of their substrates [78]. For several enzymes, the released glycosyl has been experimentally shown to retain its anomeric configuration. The active site of a GH3 enzyme has two glucosyl-binding subsites, designated as −1 and +1; the junction of these two subsites is the location where the enzyme nucleophile and general acid/base residue are found [79]. Detailed studies to further comprehend the kinetics and mechanism of the GH3 enzymes have been carried out encompassing the β-glucosidases from the Gram-negative bacteria *Thermotoga neapolitana* [80], GH3 glycosylceramidase from the Gram-positive bacteria *Paenibacillus* sp. TS12 [81] and fungi, *Aspergillus wentii* [82], *Flavobacterium meningosepticum* [83] and *Aspergillus niger* [84,85]. The kinetics of the ‘bifunctional’ β-D-glucan glucohydrolases and α-L-arabinofuranosidase/β-D-xylopyranosidases from barley have been described [86]. Similarly, kinetic and mechanistic analyses of N-acetyl-β-D-glucosaminidases from Gram-negative microbes such as *Salmonella typhimurium* [87], *Vibrio cholerae* [88] and *Vibrio furnissi* [89] as well as the Gram-positive *Bacillus subtilis* [90] also corroborate reports describing the catalytic nucleophile of the GH3 enzymes as being conserved, while the location and identity of the general acid/base residue as non-conserved.

**GH family 5**

GH family 5 (GH5), formerly known as cellulase family A, is a huge GH family that belongs to Clan GH-A. This family includes an array of enzymes found only in prokaryotes, eukaryotes and viruses, but not in humans. In fact, more than 3000 GHs enzyme sequences have been successfully identified in the CAZy database. The current 51 sub-families were grouped from 80% of the known sequences (GH5-1 to GH5-53), excluding GH5-3 and GH5-6, which have been merged into GH5-4 and GH5-5, respectively [91].

To date, there are 51 reported GH5 three-dimensional structures. The enzymes of GH5 consist of an amino-acid chain which forms a β(α)β fold, creating an open groove surrounding a conserved active site which harbours the catalytic nucleophile Glu and acid/base Glu at the C-terminus of β-strand 7 and β-strand 4, respectively. During catalysis, the carbohydrate substrate binds to the substrate-binding site from the non-reducing end (−subsites) to the reducing end (+ subsites). Typically, the GH5 enzymes have a conserved amino-acid residue (glutamic acid), which is also the catalytic residue [92].

**GH family 9**

It is the second largest cellulase family and encompasses mainly endoglucanases with a small number of processive ones [92]. Pertinently, the processive endoglucanases all contain a CBM of the 3C family naturally strongly attached to the C-terminus of the catalytic domain [93]. The cellulases in the GH family 9 (GH9) are...
mainly divided into two major sub-groups, namely, the El and Ell. The former contains only bacterial cellulases, of both aerobes and anaerobes, while the latter comprises cellulases of bacterial and non-bacterial origin [94]. All common plant cellulases are grouped under GH9, while the remaining members include cellulases that are of eubacterial, archaeal, arthropod, Echinodermata, earthworm, chordate and mollusk origin. Characteristically, the GH9 endoglucanases display appreciable activity on soluble cellulose derivatives such as carboxymethylcellulose, plant polysaccharides and phosphoric acid swollen non-crystalline cellulose, but little or no activity on crystalline cellulose [95,96].

The mechanism by which the GH9 enzymes catalysis is achieved is by the inversion of anomeric stereochemistry [92]. The catalysis is mediated by three amino acids that form the catalytic triad of the GH9 enzymes, consisting of a conserved Glu residue as the general catalytic acid and two Asp residues. One of the Asp residues, E424, functions to bind the catalytic water, while two Glu residues, D55 and D58, act as the general catalytic bases and a Tyr residue, Y318, binds the crystalline cellulose substrates. Mutation studies confirm that the conserved Glu to Gly, Ala or Gln residues are essential in carbohydrate hydrolysis [97]. The activity of the mutant enzyme is reduced to less than 0.5% of the wild-type (WT) [97], whereas mutation of the Asp that binds the catalytic water to Asn or Ala results in reduction in enzyme activity by less than 2% of WT on all cellulose substrates [98]. It is known that all catalytic domain structures of GH9 have an (a/b)_6 barrel fold that features an open active site groove consisting of at least six sugar-binding subsites —4 to +2 [99].

**Lytic polysaccharide monoxygenase (LPMO)**

Classical cellulases form a major part of the enzymes in different glycoside hydrolase (GH) families, which hydrolyse the glycosidic bonds in glucose polymers. However, the conventional hydrolytic model of cellulose polymerization has been challenged for the past few years. There have been questions on the occurrence and functional relevance of a novel class of glycolytic oxidative enzymes of both fungal and bacterial origin [100,101]. Several studies have shown that the secreted enzymes are capable of catalysing the cleavage of glycosidic bonds of glucose polymers through oxidative mechanism instead of the hydrolytic route [102–104]. These enzymes have since been reclassified as LPMOs [105], as their oxidation reaction for cleaving the glycosidic bonds is copper-dependent, producing oxidized chain-ends of either aldonolactone or a 4-ketoaldehyde [106–109]. The requirements for such reactions include molecular oxygen and an extracellular electron source, which could be supplied by celllobiose dehydrogenase (CDH) or small molecule reductants present in the lignocellulosic biomass. This fundamentally unique mechanism of cellulose chain-cleavage is assumed to circumvent the energetically difficult removal of glucan chain from highly crystalline cellulose, thereby creating new accessible ends for exoglucanase action. The oxidized carbon position in the glucan chain can also vary, as some LPMOs solely act at positions C1 or C4, and a third group oxidizes either the C1 or C4 sites [110–114].

**The discovery of LPMOs**

The first fungal LPMOs were identified as secreted enzymes that degrade cellulose, during bioprospecting works carried out in the early 1990s [115,116]. At the beginning, these enzymes were reported as hydrolyses and were classified as family 61 glycoside hydrolase (GH61) enzymes until late 2011. These enzymes were named PMOs (polysaccharide monoxygenases), and later, LPMOs within the auxiliary activity families [104,107,117–119].

In 2001, a cellulase TrCel61 produced by *T. reesei* was reported to have four types of endoglucanases that showed hydrolytic activity on cellulose [120], although this activity was hundred-fold lower than that of other known *T. reesei* endoglucanases. It was only later that researchers found that the very low cellulolytic activity of TrCel61 on all polysaccharide substrates was due to contamination. In 2008, the first crystal structure of the TrCel61 cellulase showed that the enzyme protein has a highly conserved flat surface, unlike the tunnel or cleft active sites typically found in cellulases. The TrCel61 cellulase also noticeably lacked the conserved carboxylate residues that catalyse hydrolysis. The enzyme does show weak structural similarities with CBP21, a chitin-binding protein (CBP) from the bacterium *Serratia marcescens* that is believed to improve the efficiency of the bacteria to degrade chitin. Chitin is another example of crystalline polysaccharide, similar to cellulose, consisting of β-1,4-linked N-acetylglucosamine [121,122].

Expression and secretion of GH61 in response to cellulose was initially observed in *T. reesei* [123] and subsequently, in some other fungi [124–127]. A surprising number of GH61 genes in some fungal genomes have been found, as many cellulolytic fungal species have significantly more genes for GH61 than for cellulases. An in-depth biochemical characterization of GH61s by Harris et al. [106] showed that the activity of cellulases on acid-pretreated corn stover could be enhanced, but not on pure cellulose, using an unknown mechanism. Later in 2010, the bacterial CBP was reported to be the enzyme responsible for catalysing the oxidative degradation of
chitin in the presence of molecular oxygen and a chemical reductant [102]. The chitinolytic activity was reported to be attributed to Mg$^{2+}$ and Zn$^{2+}$ metal ions that cannot generate an oxidant from molecular oxygen. A similar reaction has also been proposed to occur in fungal GH61 cells [102].

Soon findings from several major experiments linked the bacterial CBP21 oxidative cleavage reaction to the GH61 fungal enzymes. This reaction occurs through extracellular electron sources that reductively activate the GH61s [103,107,108]. This was likely due to most fungi expressing CDH. It is an extracellular hemflavoenzyme from the glucose-methanol-choline oxidoreductase superfamilly that catalyses the oxidation of cellobiose to celllobionolactone [128,129]. Langston et al. [103] suggested that the CDHs were important in activating functions of the GH61s. Consequently, a genetic study proved this theory by deleting the key CDH isoform in Neurospora crassa to result in a twofold decrease in secreted cellulase activity on pure cellulose substrate in the mutant enzyme [107].

The mechanism of enzyme activation was thought to begin with copper as the functional active site metal with the products of GH61s catalysis being generated at both the reducing and non-reducing ends of the glucan chain. Initially, the crystal structures of GH61 and bacterial family 33 carbohydrate-binding modules (CBM33) were solved with nickel [130], magnesium and zinc [106] in the binding site. In 2011, Quinlan et al. [108] solved the first crystal structure of Thermoascus aurantiacus GH61A (TaGH61A) and confirmed that the GH61 was bound to copper. Their finding was consistent with a report that natively purified N. crassa GH61s contained copper and only copper facilitated the catalysis of GH61 INCU01050 secreted by N. crassa [107]. Most importantly, the two studies provided evidence that GH6 could oxidize the non-reducing end of sugars. In contrast, studies by Quilan et al. [108] and Philips et al. [107] indicated that GH61 enzymes could also oxidize sugars at positions C6 and C4, respectively. In 2012, Beeson et al. [110] presented experimental evidence for C4 oxidation and, following the observed common catalysed reactions, they suggested the name of the enzymes from GH61s and CBM33s to be changed to PMOs. They also suggested classifying POMs as type 1 and type 2 PMOs based on their ability to oxidize both the reducing and non-reducing ends of the glucan chain.

**LPMO structure**

The first indication that GH61 enzymes were not glycoside hydrolases came with the establishment of the GH61 crystal structure of TrCel61B in 2008 [130]. The structure of TrCel61B was found to lack the identifying active site cleft of the acid/base residues, unlike those seen in glycoside hydrolases. Instead, the active site of the enzyme was revealed to be a flat surface with a supposed metal-binding site. Afterwards, structural studies on fungal cellulose-active LPMOs between 2010 and 2013 shed more light on the LPMO function. These studies focused on the active site residues internal electron transfer, substrate-binding interactions and the regioselectivity of oxidation [106,108,111,131]. Similar structural studies on bacterial LPMOs, especially on the chitin-active ones have further contributed to the present knowledge on the active site of such enzymes [111,121,131]. So far, most structural studies on LPMO have utilized X-ray crystallography, while the structures of CBP21 were determined using nuclear magnetic resonance (NMR) [132], and in some cases assisted with computational software to further enhance structural analyses [111,131]. Advancements in protein analysis technologies have resulted in successful visualization and resolution of structures from a new family of chitin-active fungal LPMOs [118] and bacterial LPMOs with cellulose-related activity [104].

**Copper-catalysed monooxygenase activity**

Vaaje-Kolstad et al. [102] first reported the oxygen- and reductant-dependent oxidative activity from the chitinolytic bacterium S. marcescens, which secretes CBP21. The oxidative activity of this enzyme was formerly accredited to a redox divalent metal ion like Mg$^{2+}$ or Zn$^{2+}$ in the active site. Although the metal ions were not associated to any oxidative activity, prior reports have indicated that divalent metal ions are required for stimulating cellulose breakdown by some GH61 proteins [106]. The finding of that study provided the first clear explanation on the chemistry of LPMOs. Subsequently, studies on fungal cellulose-active LPMOs involving reconstructed apo-LPMOS with different metal ions further revealed that copper is the native metal co-factor of LPMOs [102]. As a matter of fact, copper was later found to exist in chitin- and cellulose-active bacterial LPMOs [104,105,118,131–133] and chitin-active fungal LPMOs [118].

While there were difficulties in analysing the reaction products, which made determination of the substrate oxidation site in cellulases challenging, several researchers managed to identify key oxidation sites. Vaaje-Kolstad et al. [102], using oxygen atoms labelled with isotopes either as $H_2^{18}O$ or $^{18}O_2$, showed that the C1 position of the N-acetyl-glucosamine unit in chitin is oxidized to the corresponding carboxylic acid. A similar observation was later described in other cellulose-active bacterial and fungal LPMOs [104,107,108,110,113,114,131,134–136], while
certain cellulases were found to oxidize C4 and/or specifically oxidize only the C4 position of cellulose.

In view of such findings, a mechanism for the oxidative cleavage of glycosidic bonds was proposed (Figure 3). In this mechanism, the LPMOs attach to the hydroxyl groups of either the C1 or C4 of the glycosidic bond in cellulose to form unstable hemiketal intermediates. These intermediates undergo elimination to produce either aldonolactones (C1 oxidation) or 4-ketoaldoses (C4 oxidation), in which the former undergo either spontaneous or enzyme-catalysed hydrolysis to form aldonic acid products [137]. A combination of mass spectrometry and chemical derivation [111] ascertained that 4-ketoaldoses were the products of C4 oxidation. This finding was later confirmed by Isaksen et al. [113] using two-dimensional NMR. Similar studies using mass spectrometric analyses of reaction products suggested that oxidation at C6 by LPMOs was possible [107,135]. However, Vu et al. [114], who studied the regioselectivity of phylogenetically diverse fungal cellulose-active LPMOs, reported otherwise. Both the fungal and bacterial LPMOs could not oxidize the C6 of cellulose, presumably due to inherent differences in their active site. This discrepancy could be linked to variances in conserved catalytic residues, H-bonding and angles of orientation, inferring the use of different mechanisms by LPMOs to activate oxygen for catalysis [24].

**LPMO auxiliary activity families**

Four families of LPMOs that fall under auxiliary activity families in the CAZy database have been identified, so far [117]. These enzyme families act on cellulose as auxiliaries to the hydrolytic cellulases [102,106, 107, 108,134], thus playing a significant role in reducing the cellulase dosage for total hydrolysis [103,106]. They include: fungal AA9 LPMOs (previously GH61) that act on cellulose [106,107,110,108]; bacterial AA10 LPMOs (previously CBM33), active either on cellulose or on chitin [102,133,134]; fungal AA11 LPMOs that act on chitin [118] and fungal AA13 LPMOs that hydrolyse starch [119,136] (Figure 4). Many AA9 LPMOs act in concert with electron donors [103] such as CDHs that are co-secreted by fungi in effecting redox-mediated glycosidic bond cleavage in cellulose [103,107,138]. An interesting work by Langston et al. [103] combined a GH61 from *T. aurantiacus* with *Humicola insolens* CDH which hydrolysed cellulase to produce a mixture that contains reducing-oxidized and non-reducing end modified cellooligosaccharides [103]. The same group also demonstrated that *Thielavia terrestris* GH61 and CDH of *T. terrestris* could synergistically hydrolyse microcrystalline cellulose [103]. Interestingly, novel aldonolactonases have been discovered in the supernatant of *M. thermosthila* capable of catalysing the hydrolysis of glucono-δ-lactone, a by-product of enzymatic oxidation of cellulose.
Sugar lactones, i.e. glucono-δ-lactone, have been known to be potent inhibitors of glycosyl hydrolases [139], especially β-glucosidase [20]. The oxidative cleavage by LPMOs can also be affected by reducing agents such as glutathione or gallate ascorbic acid [102,108,109].

In line with the major role of auxiliary enzymes in cleaving glycosidic bonds, the presence of AA9 proteins is thought to enhance the activity of other cellulases [106,140] by attacking the crystalline surface on the cellulose before the action of hydrolases, creating more accessible sites for other cellulases to act [141]. Pertinently, a remarkable quality of AA9 LPMOs is the extreme expansion in the genes encoding these proteins in fungal genomes. On average, the genomes of cellulose-degrading fungi harbour as many as 10 AA9 LPMOs, and those of Aspergilli, eight ones [142–144].

Borisova et al. [145] exploited the structural basis of the unique functional properties of NcLPMO9C, a C4 oxidising AA9LPMO (LPMO9) from N. crassa, also known as NcU02916 or NcGH61-3, and found that the enzyme acts both on cellulose and on non-cellulose β-glucans such as cellobiose and xylo-β-glucan. The catalytic domain crystal structure of the NcLPMO9C revealed an expanded, high polar substrate-binding surface suitable for interaction with a variety of sugar substrates. Electron spin resonance studies showed that the Cu²⁺ centre environment in NcLPMO9C is altered upon substrate binding, although isothermal titration calorimetry analysis attributed the binding affinities in the low micromolar range for polymeric substrates, in part, to the presence of a carbohydrate-binding module (CBM1). Further comparative analysis showed that the oxidative region-selectivity of LPMO9s (C1, C4 or both) correlates with the specific structural features of the copper coordination sphere. Access to the solvent-facing axial coordination position is restricted in C1-oxidising LPMOs due to a conserved tyrosine residue, but seemingly not in C4-oxidising LPMOs. Cellulases producing a mixture of C1- and C4-oxidized products suggest adoption of an intermediate state [145].

A rather recent work by Arfi et al. [146] involving several dockerin-fused LPMOs based on enzymes from the bacterium T. fusca revealed resulting chimeras having activity levels on microcrystalline cellulose, similar to that of the WT. The complexes showed a 1.7-fold and a 2.6-fold increase in the release of soluble sugars from cellulose as compared to the free enzymes (with LPMO enhancement) and without LPMO enhancement, respectively. Hence, the suggestion that it is feasible for LPMOs to convert to the cellulosomal mode, benefitting from the proximity effects generated from the cellulosomal architecture [146].

Application of cellulases and potential application of LPMOs in industrial biofuel production

Due to the alarming level of environmental pollution through the release of green house and toxic gases from fossil fuels, the global community is now focused on biofuels, especially bioethanols. Biofuels are expected to replace 20% of the fossil fuel used by 2020. Initially, the focus was on the production of first-generation biofuels (e.g. corn bioethanol) from the bioconversion of crops such as corn grain (starch), melon seeds (fatty acids),
sugarcane (sucrose), etc. However, due to the competitiveness of these grains for food to man, cost implications as well as relative abundant availability, there has been a shift to the use of lignocellulosic biomass as feedstock for the production of bioethanol. Lignocellulose can be majorly sourced from agricultural residues such as wheat, rice, corn straw, sugarcane bagasse, to biofuels (second generation) [147,148]. There are two main processes involved in the conversion; hydrolysis of cellulose in the lignocellulosic biomass to yield reducing sugars, and fermentation of the sugars to ethanol [62]. Fungal cellulases from Trichoderma, Aspergillus and Penicillum spp. play pivotal roles in the hydrolysis process [149–153].

As is the case with most newly discovered catalysts, there will be challenges in determining the full commercial applications of LPMO enzymes. The results presented by Harris et al. [106] showed that the addition of LPMOs to (commercial) cellulase cocktails can reduce the required enzyme dose for conversion of pretreated corn stover by as much as twofold. Despite these initial auspicious results, there may be limitations in using oxygen-dependent catalysts for biomass conversion [154]. An example is the simultaneous saccharification and fermentation process for production of cellulosic ethanol that is frequently used in the production of bioethanol [155]. Under anaerobic or microaerobic conditions required for fermentation, there may be lack of oxygen for oxidative cleavage reactions catalysed by LPMOs [156,157]. If separate hydrolysis and fermentation approaches are used industrially, the LPMOs will oxidize a fraction of carbohydrate extracellularly, thus the enzymes will be no longer available for sugar fermentation. A balance between the energy losses due to oxidation by LPMOs, in feedstock, and the savings in cost from lower enzyme doses or reduced processing time must be strategically planned to maximize profits [24].

Conclusions

Cellulases occupy a central position in the degradation and efficient utilization of lignocellulosic biomass. From the previous discussion, it is clear that cellulases are not just fascinating proteins from an agricultural and industrial perspective, but are also of fundamental scientific interest. As a matter of fact, the ever-increasing demand for natural and sustainable products has further elevated the significance of these enzymes, especially in biofuel production, and has greatly changed our view of the importance of microbial cellulose degradation. In this perspective, further studies into the structure-related function of cellulases, fundamental mechanisms of their activity and protein engineering merit scientific attention. While there have been a significant accumulation of data in documenting structural features of cellulase enzyme components, research in the area of structural modelling of cellulase enzyme systems remains limited due to constrains in technological advancements for predicting protein function from structure. Hence, developing the area of functional modelling of proteins would effectively expedite the progress of informed functional models for cellulases to improve the understanding of their structure-related functions. It is a more productive means for future research into tailoring catalytic properties of various cellulases and cellulase systems for effective utilization of (ligno)cellulosic biomass for improved production of renewable chemicals and fuels.

Acknowledgments

The authors would like to acknowledge valuable help and suggestions provided by their colleagues.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Research University Grant (GUP) from Universiti Teknologi Malaysia, Johor Bahru [grant number Q.J130000.2526.13H09].

ORCID

Roswanira Abdul Wahab http://orcid.org/0000-0002-9982-6587

References


