

## A STRUCTURAL OVERVIEW OF GH61 PROTEINS – FUNGAL CELLULOSE DEGRADING POLYSACCHARIDE MONOOXYGENASES

Leila Lo Leggio<sup>a,\*</sup>, Ditte Welner<sup>b</sup>, Leonardo De Maria<sup>c</sup>

**Abstract:** Recent years have witnessed a spurt of activities in the elucidation of the molecular function of a class of proteins with great potential in biomass degradation. GH61 proteins are of fungal origin and were originally classified in family 6I of the glycoside hydrolases. From the beginning they were strongly suspected to be involved in cellulose degradation because of their expression profiles, despite very low detectable endoglucanase activities. A major breakthrough came from structure determination of the first members, establishing the presence of a divalent metal binding site and a similarity to bacterial proteins involved in chitin degradation. A second breakthrough came from the identification of cellulase boosting activity dependent on the integrity of the metal binding site. Finally very recently GH61 proteins were demonstrated to oxidatively cleave crystalline cellulose in a Cu and reductant dependant manner. This mini-review in particular focuses on the contribution that structure elucidation has made in the understanding of GH61 molecular function and reviews the currently known structures and the challenges remaining ahead for exploiting this new class of enzymes to the full.

### MINI REVIEW ARTICLE

#### Introduction

Decades of research on plant polysaccharide degrading enzymes for the exploitation of biomass have mostly focused on glycoside hydrolases, which have been classified in sequence-based families in the CAZY (Carbohydrate Active enZymes) database [1]. Glycoside hydrolases (GH) and other carbohydrate active catalytic domains are often coupled to non-catalytic carbohydrate binding modules (CBMs, reviewed in [2]), also classified in CAZY, which have the function of binding to *eg* crystalline or complex substrates and have in some cases been shown to act in synergy with the catalytic domains.

In view of the world energy crisis, bioethanol production has become a rather hot topic. While ethanol can be feasibly produced from starch rich crops, a much more renewable and sustainable solution would be the production from (ligno)cellulosic biomass, which constitutes a large proportion of agricultural and forestry byproducts. Thus a lot of attention has been devoted to enzymes able to degrade cellulose to sugars fermentable by *S. cerevisiae*. In particular the *Trichoderma reesei*/*Hypocrea jecorina* system has received much attention in terms of commercial exploitation. Cellulose breakdown (see [3] for a classic review) has been viewed for many years as carried out mainly by endoglucanases and processive exoglucanases (cellobiohydrolases) acting in synergy, often with the aid of cellulose binding domains assisting attachment to cellulose.  $\beta$ -glucosidases are also often part of cellulolytic systems, where they relieve the product inhibition of cellobiohydrolases by cellobiose, and they are often added to commercial preparations. However, the

mechanism by which some microorganisms are able to efficiently degrade crystalline cellulose has remained in many ways a mystery. In the last few years a new class of fungal proteins with huge potential for the degradation of cellulose has received much attention, the GH61 proteins. Initially classified as family 6I among the glycoside hydrolases they are now recognized to be Cu-dependent oxidases [4, 5, 6], calling for a reclassification of these enzymes. As such a reclassification is yet to be implemented in the CAZY database, we choose in this review to keep the somewhat inappropriate GH61 designation, which allows retrieval of most of the earlier literature. This family has puzzled carbohydrate active enzyme experts since its discovery, and to some extent continues to do so. Structure determination by X-ray crystallography was a crucial step towards understanding the significance and mechanism of action of these enzymes. This short review briefly summarizes the progress up to now, and focuses on the structures currently known.

#### A brief history of GH61

The first GH61 protein to be identified was probably Cell from *Agaricus bisporus* the sequence of which was described in 1992 after cloning of the gene [7]. Although no activity could be described, the gene was induced on growth on cellulose, and the presence of a sequence typical of a cellulose binding domain implicated the protein in cellulose degradation. The GH61 family was first created in 1997, when it was referred to at least twice in the literature [8, 9]. The evolution of the family in terms of number of members can be seen in Figure 1a.

The first papers on characterization of GH61 family members reported very low cellulose degrading activity if any. For example *T. reesei* Cel61A [10], showed some degrading activity on polymeric cellulosic substrates, but at levels 5-6 orders of magnitude lower than a conventional cellulase, Cel7B, making it difficult even by use of sensible controls to totally rule out the possibility of contamination by canonical cellulases. In hindsight, the low activity can be explained

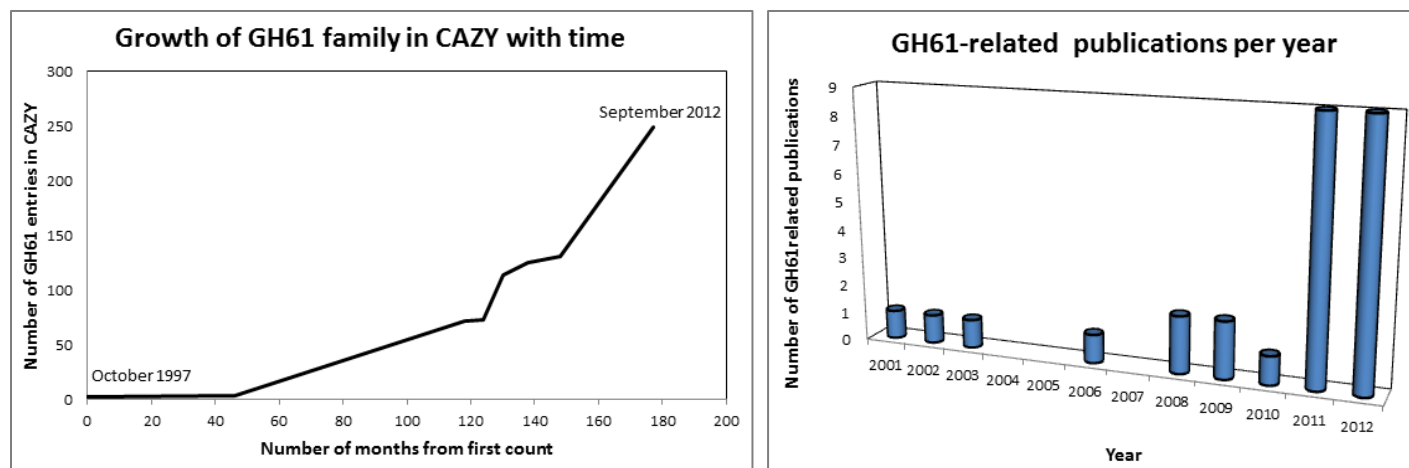
<sup>a</sup>Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100, Copenhagen Ø, Denmark

<sup>b</sup>Statens Seruminstitut, Artillerivej 5, 2300 Copenhagen S, Denmark

<sup>c</sup>Novozymes A/S, Kroghshøjvej 36, 2880 Bagsvaerd, Denmark

\* Corresponding author.

E-mail address: leila@chem.ku.dk (Leila Lo Leggio)



**Figure 1.** A) Number of GH61 members in CAZY from October 1997, when the new family was first announced in publications [8, 9]. The first count after the family was formed was in August 2001 according to [10]. Subsequent counts were made with the help of the Wayback Machine (<http://archive.org/web/web.php>). B) Number of articles in Pubmed with 'GH61' or 'family 61' in title/abstract for each year (checked for relevance). Some publications which include information on GH61 may not be included in the count, if they did not include the chosen search keywords in abstract or title.

by the lack of essential cofactors, which at the time were unknown. However the identification of GH61 members in cellulolytic organisms such as *T. reesei*, *A. bisporus*, *Aspergilli* species and *Neurospora crassa* together with their co-induction with classical cellulases upon growth on cellulose [7] [11], already early on suggested the GH61 family involvement in lignocellulose degradation. This was further supported by the fact that several of the first GH61 domains were found to be associated with family I CBMs, which are crystalline cellulose binders.

A first breakthrough in the molecular understanding of GH61 function came from the structures of two family members which were communicated at conferences and in peer-reviewed journals in 2008 [12] [13]. The first publication [13] revealed the 3D structure of *Hypocrea jecorina* (*Trichoderma reesei*) Cel61B (from now on referred to as HjGH61B), solved by Single-wavelength Anomalous Diffraction utilizing Ni ions from the crystallization conditions. HjGH61B has an immunoglobulin-like  $\beta$ -sandwich fold, and very atypically if it were a true GH, lacks a clear substrate binding groove and an appropriately positioned and exposed active site carboxylate pair, which is, with few exceptions, ubiquitous in GHs mechanisms. The authors concluded that based on the structure this was an unlikely GH. Most interestingly, a clear metal binding site was revealed in each of the two molecules in the asymmetric unit, and the coordinated metal ion was assigned as Ni due to the presence of this metal in the crystallization conditions. The metal protein ligands were the N-terminal His as well as an additional His and a Tyr, all noted to be conserved in GH61 sequences and thus of potential functional importance. However the nature or indeed a presence of the native metal could not be established in solution despite attempts to do so by particle-induced X-ray emission. Furthermore no link to a function or activity could at this point be made.

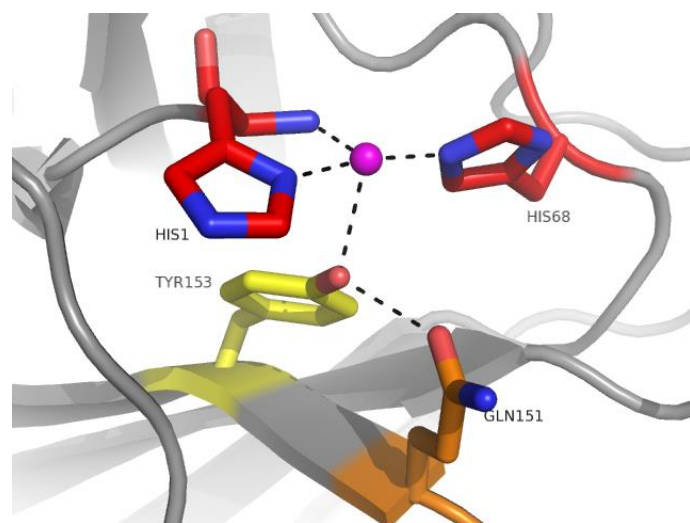
The structure of GH61E from *Thielavia terrestris* (TtGH61E), determined by Multiple Isomorphous Replacement, was preliminarily presented in 2008 [12] and published in 2010 [14]. *Thielavia terrestris* is a thermophilic cellulose degrading ascomycete, which when cultured on cellulose secretes a number of known cellulases and hemicellulases, but also at least six GH61 proteins, making up about 10% of total soluble protein [14]. TtGH61E has only 29 % sequence identity with HjGH61B, yet shares many significant structural features, notably the divalent metal binding site. As well as

structure determination of TtGH61E, Harris and coworkers present in [14] the first GH61 activity assay, previously disclosed in the patent literature [15]. The assay measured GH61 activity in terms of their boosting effect on the activity of conventional hydrolytic cellulases. It should be noted that cellulase boosting activities, which now can be assigned to GH61 enzymes, were reported already in conferences well ahead of this time, for example for a fungal extract at the 2003 MIE Bioforum [16]. The establishment of an assay, although as it turned out later a rather indirect one, was instrumental to the major breakthrough of this publication, the establishment of a firm link between the found metal binding site and cellulose degradation. By combining knowledge of the structure with an assay, the requirement of the metal for boosting activity was demonstrated by structure-guided site directed mutagenesis, showing loss of this activity in variants where the metal binding residues were removed. The His1 to Asn and His68 to Ala mutants were completely inactive, while a Tyr153 to Phe mutant had reduced activity. Mutation of Gln151 (H-bonding to Tyr153) to a variety of residues reduced activity even more severely.

In the HjCel61B and TtGH61E structures, the divalent metals were assigned purely on the basis of their presence in the crystallization mixture, following the normally good crystallographic practice of modeling only chemical entities that are known or can be proven by other means to be present in the mixture. Thus the identity of the metal binding site in its native environment remained unclear. The demonstration that GH61 could boost the activity of hydrolytic cellulases established a biological justification for the coexistence and co-expression of GH61 with the more conventional cellulose-degrading enzymes and underpinned the biotechnological potential of GH61. It is noteworthy that these first structural publications included authors from two large world enzyme producers, Novozymes A/S and Genencor. As can be seen from Figure 1b, after the publication of the TtGH61E article the literature on GH61 was significantly boosted.

Another very important link emerging from these first two structures was the structural similarity to chitin binding protein CBP21 from *Serratia marcescens* [17] at that point thought to non-enzymatically disrupt chitin and classified as a carbohydrate binding module belonging to family 33 (CBM33). Not only the overall structure of CBP21 was similar to the two GH61 structures, but

CBP2I had also a very similar arrangement of residues as in the identified metal binding site, although no divalent metal was modeled in this structure (a sodium ion is though modeled at this site in one of the molecules in the asymmetric unit, see PDB code 2BEM). Furthermore mutagenesis of the conserved non-terminal His at this site was shown to affect the boosting effect of CBP2I on chitinase C [18]. There is however an important difference at the metal binding site since CBP2I has a conserved Phe instead of Tyr, while a Tyr to Phe substitution is detrimental to the activity of TtGH61E.



**Figure 2.** Metal binding site of TtGH61E (PDB code 3EJA, chain A) highlighting the mutated residues [14]. In red residues whose replacement resulted in complete abolishment of activity. In orange Gln151 whose replacement resulted in complete abolishment or severe impairment of activity (depending on the replacement residue). In yellow Tyr153 whose replacement resulted in impairment of activity.

Given the functional and structural similarities and their complementary phylogenetic distribution (GH61 are predominantly fungal – see also the phylogenetic analyses in [14] and [19] – while CBM33 are predominantly bacterial and viral), it was already suspected at this time that the two families maybe distantly evolutionarily related. More details on the CBM33 family can be found in a recent publication which reviews the two families' (CBM33 and GH61) biotechnological potential [20].

Although this review strictly focuses on GH61, the progress in the understanding of the two families has been so linked that CBM33 cannot be completely ignored. A key paper in the understanding of the mechanism showed that CBM33 proteins oxidatively degrade crystalline chitin [21] producing a mixture of oxidized and unoxidized even-numbered chitoooligosaccharides, preferentially of high DP. One of the oxygens in the oxidized products came from water while the other was contributed by molecular oxygen. Furthermore this publication established that a divalent metal ion was necessary for oxidative function although the nature of the metal requirement was not firmly established. It was later shown that CBM33s are not exclusive to chitin degradation but are also able to degrade cellulose, since Cels2 from *S. coelicolor* A3(2) was shown to degrade crystalline cellulose with production of oxidized (aldonic acids) and unoxidized products, again with dominance of even-numbered DP [22]. Similarly to previous work, divalent metal ions were shown to be necessary, but no clear preference could be shown.

Not long after researchers began to report oxidative activity also for GH61 proteins, resulting in a mixture of non-oxidized and oxidized cellodextrins after crystalline cellulose degradation [4, 23, 5]. In [4], it was clearly shown that at pH 5 a GH61 protein from *Thermoascus aurantiacus*, TaGH61A, is highly selective for binding of Cu<sup>2+</sup> ions and that the Cu-loaded TaGH61A oxidatively degrades crystalline cellulose in the presence of small molecule redox active agents such as gallate and ascorbate. Electron paramagnetic resonance spectroscopy showed clearly a signal for Cu(II) similar to the one observed in type II copper oxygenases. More or less at the same time [24], it was also shown that the combination of GH61 and cellobiose dehydrogenases (CDH) from same or different organisms could oxidatively degrade highly crystalline cellulose without added small molecule reductants. Later in 2011, reports in [5] and [6] confirmed that GH61 are cellulose degrading Cu metalloenzymes. In [5] the important role of cellobiose dehydrogenases (CDH) in cellulose degradation by GH61 was underpinned by genetic and biochemical experiments, and it was suggested that in nature reduced CDHs may reduce Cu(II) to Cu(I) in the catalytic mechanism of GH61s. Furthermore it was suggested that polysaccharide monooxygenases, as GH61 are referred to in this publication, can be of two types depending on whether oxidation is introduced on one side or the other of the broken glycosidic bond. The mechanisms of the type I and type 2 GH61s have been investigated in more detail in [25] and by isotope labelling GH61s of type I were shown to incorporate one oxygen atom from molecular oxygen into the product and therefore to be monooxygenases. Two crystal structures for *Neurospora crassa* GH61s have recently been published [26], where the authors claim to have isolated dioxygen species, which however are difficult to unequivocally establish purely by crystallography, despite high resolution and careful refinement. As a final remark, after the very recent elucidation of the chitinolytic system of *Enterococcus faecalis* V583 and structure determination of a new CBM33 enzyme, Cu has after some dispute been recognized to be the active metal also for these enzymes [27]. Thus there seems to be general consensus that both GH61 and CBM33 are Cu dependent monooxygenases.

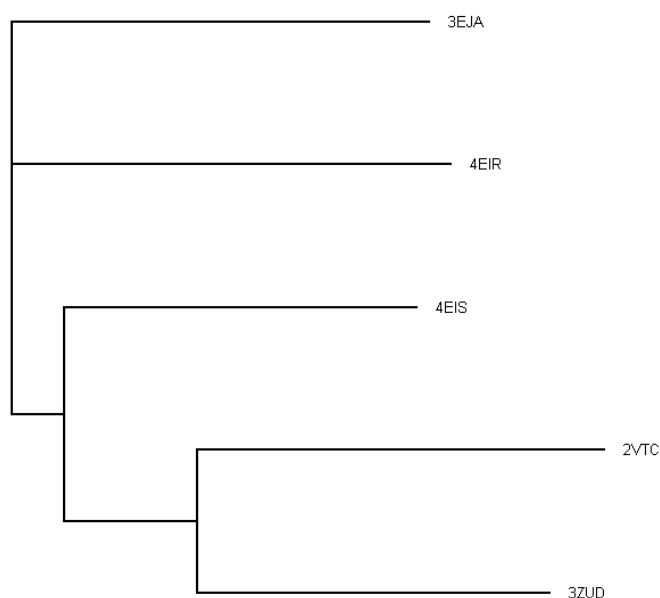
Structural biology has thus made an essential contribution to the understanding of the molecular mechanism of GH61 function, especially by the discovery of the similarity between GH61 and CBM33 and the metal binding site, its functional importance (by guiding mutagenesis) and its clarification as a Cu(II) binding site. The structures known so far for GH61 and some of their features are reviewed below.

### Known structures of GH61 family members

To date, five structures of GH61 proteins have been determined, all by X-ray crystallography. Table 1 summarizes some of the characteristics of the structures, while table 2 summarizes the sequence and structural similarity between them. Two of the structures (TtGH61E and HjGH61B) have already been discussed in some detail above and are in fact the two most dissimilar (Dali-Lite [28] aligns 200 residues with 1.9 Å Cα rmsd and 29% structure based sequence identity). Three additional structures have since been determined [4] [26]. The dendrogram in Figure 3 illustrates graphically the relationship between the different proteins. TtGH61E (code 3EJA) and NcPMO2 (code 4EIR) are distant from each other and the other sequences, while NcPMO3 (code 4EIS), HjGH61A (code 2VTC) and TaGH61A (code 3ZUD) form a more closely related group. This is also highlighted in Figure 4, where differences in loop structures as well as presumed functional residues are clearly visible between TtGH61E, NcPMO2 and NcPMO3.

**Table 1.** Overview of known structures. The Tyr conserved at the aromatic surface in all structures is in bold.

Abbreviation	Organism	PDB code(s)	Resolution (Å)	Associated with CBM1	'Flat' surface aromatics	Ref
TtGH61E	<i>Thielavia terrestris</i>	3EII	2.25	No	Tyr67, <b>Tyr191</b> , Tyr192	[14]
		3EJA	1.90			
NcPMO2	<i>Neurospora crassa</i>	4EIR	1.10	No	Tyr67, <b>Tyr206</b> , Trp207	[26]
TaGH61A	<i>Thermoascus aurantiacus</i>	3ZUD	1.25	No	Tyr24, <b>Tyr212</b>	[4]
		2YET	1.50			
HjGH61B	<i>Hypocrea jecorina (Trichoderma reesei)</i>	2VTC	1.60	No	Tyr23, <b>Tyr212</b>	[13]
					Tyr20, Tyr24, Tyr163, <b>Tyr210</b>	
NcPMO3	<i>Neurospora crassa</i>	4EIS	1.37	No	Tyr24, Tyr163, <b>Tyr210</b>	[26]



**Figure 3.** Dendrogram (produced by near joining method in ClustalW2 [29] and displayed with Treeview [30]) graphically showing the similarities between the proteins of known structure, indicated by their PDB codes (the proteins corresponding to each PDB code can be found in Table 1). The multiple structural alignments were carried out with the server version of Mammoth-Mult [31].

### Association of GH61 catalytic domain with Carbohydrate Binding Modules (CBMs)

As noted also in the early reports on GH61 gene cloning and sequences, these catalytic domains are often associated with CBMs [2], and in particular with CBM1. In [14] about 20% of GH61 sequences were estimated to be associated with a C-terminal CBM1 (an N-terminal CBM would interfere with the N-terminal His metal-binding function). CBM1s are A-type CBM, typically presenting a flat surface which binds to a crystalline polysaccharide, for CBM1 usually cellulose [2]. A recent search using the Cazymes Analysis Toolkit (CAT [32]) shows that 37 out of 143 (26%) of GH61s in CAT are

associated with CBM1, indicating that the estimate in [14] is holding up as new sequences come into CAZY. It seems that in organisms having multiple GH61 genes some are associated with a CBM1 and some are not, for example in *Heterobasidion irregulare*, three out of ten GH61 genes have an associated CBM1 sequence [33]. Two of the CAZY entries are interestingly associated with CBM18s, which are typically chitin fragment binding. This could perhaps suggest that like some CBM33s can degrade cellulose, some GH61 could be involved in chitin degradation. CBM18s are however normally considered type C CBMs, binding small fragments rather than crystalline polysaccharides. Interestingly, the CBM33 CelS2 protein shown to be active on cellulose [22] has a CBM2 associated with it (CBM2s are A-type binders usually binding to cellulose, but also to chitin or xylan).

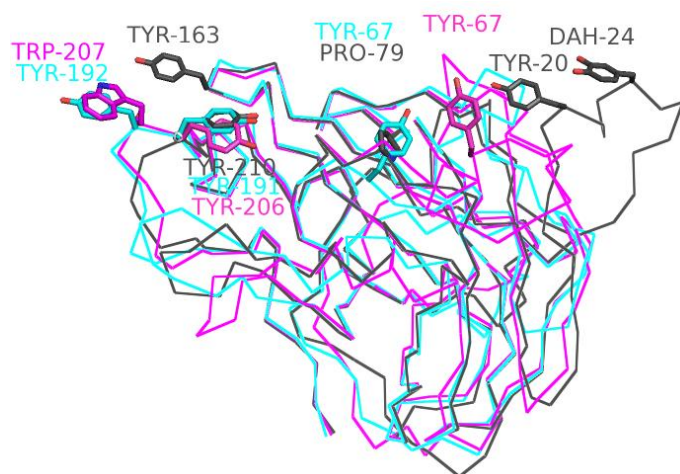
None of the structurally characterized GH61 proteins has naturally a CBM attached. This is not surprising as successful crystallization is strongly biased towards single domain, compact proteins.

However for the TtGH61E structure [14], a CBM1-like feature within the catalytic domain was noted, three Tyr forming a flat surface and arranged similarly as the three Tyr in the structure of the CBM1 of *T. reesei* Cellobiohydrolase I [34]. For TtGH61E, crystalline cellulose binding activity has been qualitatively shown experimentally [35] and it has been shown that substitution of one of the aromatics (Tyr192) to Ala reduces activity [14]. Although this CBM1-like feature is not generally conserved, all of the structures of GH61 determined so far have a 'flat' face, to which two or more aromatic residues contribute (see Table 1 and Figure 4 for illustration of three of the most diverse structures) which could be involved in crystalline substrate binding. Tyr191 of TtGH61E has an equivalent Tyr in all other structures determined (in bold in Table I). A structural equivalent of Tyr192 is present as Trp207 in NcPMO2, otherwise the loops differ here between the structures. The third Tyr forming the flat surface in TtGH61E is often but not always a Pro in other structures (HjGH61A, TaGH61A and NcPMO3). NcPMO3 has, together with the Tyr conserved in all structures, additional aromatics Tyr20, Tyr24 and Tyr163 which form a flat aromatic surface, while NcPMO2 has a Tyr67 (which is not equivalent of Tyr67 in TtGH61E). Differences in the putative binding surfaces of GH61 have also been discussed in [26] and a flat binding surface has also been described for CBM33 [27].



**Table 2.** Structure similarity from Dali-Lite [28] (number of aligned residues,  $C\alpha$  rmsd and structure-based sequence identity are shown).

	TaGH61A (3ZUD, A)	TtGH61E (3EJA, A)	HjGH61B (2VTC, A)	NcPMO2 (4EIR, A)	NcPMO3 (4EIS, A)
TaGH61A (3ZUD, A)	-	201 res 1.7 Å 33%	220 res 1.3 Å 47%	205 res 1.7 Å 31%	218 res 1.5 Å 41%
	TtGH61E (3EJA, A)	-	200 res 1.9 Å 29%	198 res 1.7 Å 39%	199 res 1.5 Å 42%
		HjGH61B (2VTC, A)	-	205 res 1.8 Å 28%	216 res 1.7 Å 37%
			NcPMO2 (4EIR, A)	-	204 1.7 Å 42%
				NcPMO3 (4EIS, A)	-



**Figure 4.** Residues contributing to the flat potential substrate binding surface of TtGH61E (cyan), NcPMO2 (magenta) and NcPMO3 (grey). DAH is the hydroxylated form of Tyr24.

### The metal binding site

As stated above, there is currently reasonable consensus that the active metal in GH6I proteins is copper, however it seems that GH6I (and CBM33) can be sometimes demetallated/substituted for other metals during overexpression and purification despite the high affinities for Cu at active pHs. One possibility is that this is mediated by changes in pH during purification. In table 3, the distances to protein residues for the structures where metal at the active site could be positively identified as Cu (3ZUD, 4EIS, 4EIR) are shown and are very typical and very similar in the three proteins. The Cu(II) is in tetragonal coordination geometry in all structures reported. The protein ligands are highly conserved in GH6I sequences and in all the structures reported so far.

The special spatial arrangement of the two coordinating His (one coordinating both with the N-terminus and its side chain) has been named 'histidine brace' [4], and a similar arrangement has also been observed in copper methane monooxygenases, which however differ

by having two rather than one Cu atoms at a distance of 2.7 Å from each other, as confirmed by a recent 2.68 Å resolution structure [36]. Other conserved residues around the metal binding site are the Gln hydrogen bonding to the conserved Tyr and a third His residue (His164 in TaGH61A), the function of which is yet unknown. These are interestingly not the same in CBM33. In CBP2I an Asn (185) and an Asp (182) are equivalent to the GH6I Gln and His, respectively, and the two residues are highly conserved as Asn and Asp in the family. Only mutation of Asp182 to Ala affected the combined activity of CBP2I and chitinase C, while mutation of Asn185 had little effect [18]. This correlates well to the fact that the equivalent of the conserved Tyr is in CBM33 a Phe, which could not form a hydrogen bond with a polar residue. Aside from the residues directly coordinating the ligands and the additional Gln and His, there is considerable diversity in the residues immediately surrounding the Cu(II) site among the different GH6I proteins, as illustrated in Figure 5 by the structures of TtGH61E and NcPMO3. This diversity may prove important in modulating the activities of different GH6I.

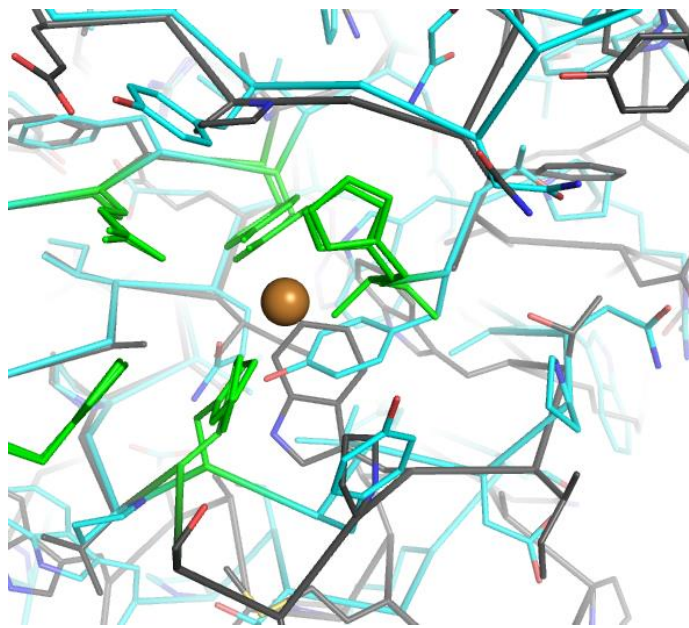
**Table 3.** Cu-protein distances in TaGH61A (3ZUD), NcPMO2 (4EIR) and NcPMO3 (4EIS). For 3ZUD the distances are to the main conformation of the Cu atom. In 4EIR and 4EIS there are two molecules per asymmetric unit, hence two distances are given.

	N-terminal His N	N-terminal His ND1	His NE2	Tyr OH
3ZUD	2.2 Å	1.9 Å	2.0 Å	2.9 Å
4EIR	2.2/2.2 Å	1.9/1.9 Å	2.0/2.0 Å	2.8/2.8 Å
4EIS	2.3/2.3 Å	1.9/1.9 Å	2.1/2.1 Å	2.7/2.8 Å

The N-terminal His plays a special role in coordinating the metal, as it provides two ligands, a main chain and a side chain nitrogen. In [4] it was first recognized that this N-terminal His is a site of unusual post-translational modification, a methylation at Nε2. This was supported by crystallographic analysis and mass spectrometry for *T. aurantiacus*. Reanalysis of previously reported structures, and

modelling of methylation in all subsequently reported structures suggests that this may be a feature of all active GH61 proteins. No evidence of such a modification has been presented for the bacterial CBM33s.

Intriguingly, although the N-terminal His is extremely well conserved in GH61, some GH61 members have an Arg at this position, for example HiGH61G from *Heterobasidion irregulare* [33], which also lacks the other Cu-coordinating His. Even more intriguingly, it seems that some of these proteins (including HiGH61G) are upregulated when fungi are grown on lignocellulosic substrates, as well as GH61s having an integral metal binding site (as judged by sequence). This observation opens the possibility that some GH61 may have additional and metal-independent roles in cellulose degradation.



**Figure 5.** Structural diversity around the Cu binding site. TtGH61E is shown in cyan, while NcPMO3 in grey. The conserved Cu-binding residues and additional His and Gln are in green.

## Outlook

One of the areas of interest in terms of exploiting the biotechnological potential is of course discovery of novel GH61 enzymes. In this sense it seems that genome mining and generally 'omics' analyses in lignocellulose degrading organisms may prove to be a very fruitful strategy for GH61 as well as other plant cell wall degrading enzymes. Especially white-rot fungi such as *Phanerochaete chrysosporium* [19], thermophilic biomass degrading fungi where the GH61 family is largely expanded – eg. *Thielavia terrestris*, boasting 18 GH61 genes compared to 3 in *T. reesei* [37] – and plant pathogens [38] have been subject of great attention. In the last few years, upregulation of some, but not all, GH61s upon growth on cellulosic substrates has been observed in transcriptome and secretome analyses of *P. chrysosporium* [19], transcriptome analysis of *Phanerochaete carnosus* [39], proteomic analysis of *Aspergillus nidulans* growing on sorghum stover [40], and qRT-PCR studies on GH61 of the pathogen *Heterobasidion irregulare* [33], where HiGH61H showed a rather spectacular 17,000 fold increase on spruce heartwood. Since some GH61 genes/proteins are not

upregulated by growth on cellulosic substrates, these may have different substrate specificities, which are yet to be explored.

The interplay between CDH and GH61 needs to be explored further. Co-induction of CDH and GH61 upon growth on cellulosic substrates has been reported in several large scale studies and organisms, among others *P. chrysosporium* [19], *Aspergillus nidulans* [40] and *Thielavia terrestris* [24, 23]. It has also been suggested that lignin may be sufficient as a reductant, as no addition of small molecules reductants is needed when lignin is used [41] [42].

Although oxidative enzymes like GH61 and CBM33 act in synergy with glycoside hydrolases, the final oxidised products of the reaction can pose a limit to the final yields that can be obtained. Gluconic acid, which can be a significant proportion of overall reaction products from commercial enzyme preparations [41], is a known inhibitor [43] of  $\beta$ -glucosidase and more inhibitory to  $\beta$ -glucosidase activity than glucose in realistic reaction mixtures [41]. Furthermore it is not fermentable by *S. cerevisiae*. Cellobionic acid is also a worse substrate than cellobiose for  $\beta$ -glucosidases [41]. A better understanding of the interplay of different components is necessary, to achieve as high as possible yields of conversion.

Many challenges remain also on the fundamental understanding of GH61 action, their interaction with substrate and the electron transfer pathways. Crystallography is expected to continue contributing to the elucidation of GH61 detailed function and diversity. However a true molecular understanding will not come from structural biology alone, but requires combined efforts involving experts from different fields including phylogenetics, transcriptomics and bioinorganic chemistry.

## Acknowledgements

LLL and DW thank Novozymes A/S for the long standing collaboration on GH61 proteins and provision of proteins for structural analysis, in particular in Denmark Kirk Schnorr, Katja Solomon Johansen and Morten Tovborg, and in the US Paul Harris, Matt Sweeney and lastly Feng Xu for critical reading of this manuscript. LLL thanks also Jens-Christian N. Poulsen at the University of Copenhagen and her former students Harm Otten and Veronika Karlsson for their involvement in the project. Furthermore LLL thanks Professors Gideon Davies, Paul Walton and Bernard Henrissat for many fruitful discussions on GH61 proteins.

## Citation

Lo Leggio L, Welner D, Maria LD (2012) A structural overview of GH61 proteins – fungal cellulose degrading polysaccharide monooxygenases. *Computational and Structural Biotechnology Journal*. 2 (3): e201209019.  
doi: <http://dx.doi.org/10.5936/csbj.201209019>

## References

1. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nuc Ac Res* 37:D233-D238
2. Boraston AB, Bolam DN, Gilbert HJ and Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382:769-781
3. Béguin P (1990) *Molecular Biology of Cellulose Degradation*. *Ann Rev Microbiol* 44:219-248

4. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen J-CN et al (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. PNAS 108:5079-15084
5. Phillips CM, Beeson WT, Cate JH and Marletta MA (2011) Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. ACS Chemical Biology 6:1399-1406
6. Westereng B, Ishida T, Vaaje-Kolstad G, Wu M, Eijsink VGH et al (2011) The putative endoglucanase PcGH61D from *Phanerochaete chrysosporium* is a metal-dependent oxidative enzyme that cleaves cellulose. PLOS One 6:e27807
7. Raguz S, Yagiie E, Wood D and Thurston C (1992) Isolation and characterization of a cellulose-growth-specific gene from *Agaricus bisporus*. Gene 119:183-190, 1992.
8. Saloheimo M, Nakari-Setälä T, Tenkanen M and Penttilä M (1997) cDNA cloning of a *Trichoderma reesei* cellulase and demonstration off endoglucanase activity expression in yeast. Eur J Biochem 249:584-591
9. Henrissat B and Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol 7:637-644.
10. Karlsson J, Saloheimo M, Siika-aho M, Tenkanen M, Penttilä M et al (2001) Homologous expression and characterization of Cel61A (EG IV) of *Trichoderma reesei*. Eur J Biochem 268:6498-6507
11. Wymelenberg AV, Denman S, Dietrich D, Bassett J, Yu X et al (2002) Transcript Analysis of Genes Encoding a Family 61 Endoglucanase and a Putative Membrane-Anchored Family 9 Glycosyl Hydrolase from *Phanerochaete chrysosporium*. App Env Microbiol 68:5765-5768
12. D. Welner, M. H. Jensen, K. McFarland, J.-C. N. Poulsen, H. Otten et al (2009) Structure and function of enzymes in biomass degradation. In *Biotechnology of lignocellulose degradation and Biomass Utilization - Mie Bioforum 2008*
13. Karkehabadi S, Hansson H, Kim S, Piens K, Mitchinson C et al (2008) The first structure of a glycoside hydrolase family 61 member, Cel61B from *Hypocrea jecorina* at 1.6 Å resolution. J Mol Biol 383:144-154
14. Harris PV, Welner D, McFarland KC, Re E, Poulsen J-CN et al (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: Structure and function of a large, enigmatic family. Biochemistry 49:3305-3316
15. Brown K, Harris P, Zaretsky E, Re E, Vlasenko E et al (2005) Polypeptides having cellulolytic enhancing activity and polynucleotides encoding the same. Patent WO/2005/074647.
16. Xu F (2004) Enhancing biomass conversion to fermentable sugars: A progress report of a joint government-industrial project. In *Biotechnology of Lignocellulose Degradation and Biomass Utilization - MIE Bioforum 2003*.
17. Vaaje-Kolstad G, Houston D, Riemen A, Eijsink V and van Aalten D (2005) Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. J Biol Chem 280: 11313-11319
18. Vaaje-Kolstad G, Horn SJ, van Aalten DMF, Synstad B and Eijsink VGH (2005) The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. J Biol Chem 280:28492-28497
19. Wymelenberg AV, Gaskell J, Mozuch M, Kersten P, Sabat G et al (2009) Transcriptome and secretome analyses of *Phanerochaete chrysosporium* reveal complex patterns of gene expression. App Env Microbiol 75:4058-4068
20. Horn SJ, Vaaje-Kolstad G, Westereng B and Eijsink VG (2012) Novel enzymes for the degradation of cellulose. Biotech Biofuels 5:45
21. Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H et al (2010) An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides Science 330:219-222
22. Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunæs AC, Stenström Y et al (2011) Cleavage of cellulose by a CBM33 protein. Prot Sci 20:1479-1483
23. Langston JA, Shaghasi T, Abbate E, Xu F, Vlasenko E et al (2011) Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. App Env Microbiol 77:7007-7015
24. Langston JA, Brown K, Xu F, Borch K, Garner A and Sweeney MD (2012) Cloning, expression, and characterization of a cellobiose dehydrogenase from *Thielavia terrestris* induced under cellulose growth conditions. BBA 1824:802-812
25. Beeson WT, Phillips CM, Cate JHD and Marletta MA (2012) Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. JACS 134:890-892
26. Li X, Beeson WT, Phillips CM, Marletta MA and Cate JH (2012) Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. Structure 20:1051-1061
27. Vaaje-Kolstad G, Bøhle AL, Gåseidnes S, Dalhus B, Bjørås M et al (2012) Characterization of the chitinolytic machinery of *Enterococcus faecalis* V583 and high-resolution structure of its oxidative CBM33 Enzyme. J Mol Biol 416:239-254
28. Holm L, Kääriäinen S, Wilton C and Plewczynski D (2006) Using Dali for structural comparison of proteins. Curr Prot Bioinf Chapter 5:5.5.1-5.5.24
29. Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P et al (2007) ClustalW and ClustalX version 2. Bioinformatics 23:2947-2948
30. Page, RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers 12:357-358
31. Lupyan D, Leo-Macias A and Ortiz A (2005) A new progressive-iterative algorithm for multiple structure alignment. Bioinformatics 21:3255-3263
32. Park BH, Karpinets TV, Syed MH, Leuze MR and Uberbacher EC (2010) 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
33. Yakovlev I, Vaaje-Kolstad G, Hietala AM, Stefańczyk E, Solheim H et al (2012) Substrate-specific transcription of the enigmatic GH61 family of the pathogenic white-rot fungus *Heterobasidion irregulare* during growth on lignocellulose. App Micr Biotech 95:979-990
34. Kraulis P, Clore G, Nilges M, Jones T, Pettersson G et al (1989) Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. Biochemistry 28:7241-7257
35. Welner D (2006) Four proteins in glycobiology - structure and function. MsC thesis, University of Copenhagen.
36. Smith S, Rawat S, Telser J, Hoffman B, Stemmler T et al. (2011) Crystal structure and characterization of particulate methane monooxygenase from *Methylocystis* species strain M. Biochemistry 50:10231-10240
37. Berka RM, Grigoriev IV, Otilar R, Salamov A, Grimwood J et al. (2011) Comparative genomic analysis of the thermophilic biomass degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. Nature biotech 29:922-927

**What is the advantage to you of publishing in *Computational and Structural Biotechnology Journal (CSBJ)* ?**

- ✚ Easy 5 step online submission system & online manuscript tracking
- ✚ Fastest turnaround time with thorough peer review
- ✚ Inclusion in scholarly databases
- ✚ Low Article Processing Charges
- ✚ Author Copyright
- ✚ Open access, available to anyone in the world to download for free

[WWW.CSBJ.ORG](http://WWW.CSBJ.ORG)

38. Gibson DM, King BC, Hayes ML and Bergstrom GC (2011) Plant pathogens as a source of diverse enzymes for lignocellulose digestion. *Curr Opin Microbiol* 14:264-270
39. MacDonald J, Doering M, Canam T, Gong Y, Guttman DS et al (2011) Transcriptomic responses of the softwood-degrading white-rot fungus *Phanerochaete carmosa* during growth on coniferous and deciduous wood. *App Env Microbiol* 77:3211-3218
40. Saykhedkar S, Ray A, Ayoubi-Canaan P, Hartson SD, Prade R et al (2012) A time course analysis of the extracellular proteome of *Aspergillus nidulans* growing on sorghum stover. *Biotec Biofuels* 5:52

41. Cannella D, Hsieh C-WC, Felby C and Jørgensen H (2012) Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content. *Biotec Biofuels* 5:26
42. Dimarogona M, Topakas E, Olsson L and Christakopoulos P (2011) Lignin boosts the cellulase performance of a GH-61 enzyme from *Sporotrichum thermophile*. *Biores Techn*, 110:480-487
43. Dale M, Ensley H, Kern K, Sastry K and Byers L (1985) Reversible inhibitors of  $\beta$ -glucosidase. *Biochemistry* 24:3530-3539

**Competing Interests:**

LEDM is an employee of Novozymes A/S, a major enzyme producer.



© 2012 Lo Leggio et al.

Licensee: Computational and Structural Biotechnology Journal.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are properly cited.