Carbohydrate active enzyme domains from extreme thermophiles –

components of a modular toolbox for lignocellulose degradation

Jonathan Botha[1,2,4], Eshchar Mizrachi[2,3], Alexander A. Myburg[2,3], Don Cowan[1,2,4]*

*Corresponding author: Don.Cowan@up.ac.za

[1]Centre for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, Private Bag X20, Pretoria, 0028, South Africa

[2]Department of Genetics, University of Pretoria, Private Bag X20, Pretoria, 0028, South Africa

[3]Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, 0028, South Africa

[4]Genomics Research Institute (GRI), University of Pretoria, Private Bag X20, Pretoria, 0028, South Africa

Abstract

Lignocellulosic biomass is a promising feedstock for the manufacture of biodegradable and renewable bioproducts. However, the complex lignocellulosic polymeric structure of woody tissue is difficult to access without extensive industrial pre-treatment. Enzyme processing of partly depolymerised biomass is an establishedtechnology, and there is evidence that high temperature (extremely thermophilic) lignocellulose degrading enzymes (CAZymes) may enhance processing efficiency. However, wild-type thermophilic CAZymes will not necessarily be functionally optimal under industrial pretreatment conditions. With recent advances in synthetic biology, it is now potentially possible to build CAZyme constructss from individual protein domains, tailored to the conditions of specific industrial processes.. In this review, we identify a 'toolbox' of thermostable CAZyme domains from extremely thermophilic organisms and highlight recent advances in CAZyme engineering which will allow for the rational design of CAZymes tailored to specific aspects of lignocellulose digestion.

Keywords

Lignocellulose, CAZyme, Extreme thermophiles, Synthetic biology, Protein domains

Introduction

Biomaterials (materials derived from renewable and sustainable biological substrates) could provide an economically viable strategy to mitigate or ameliorate environmental challenges and potentially replace products derived from petrochemical feedstocks (Naik et al. 2010). However, to maximise efficiency of bioproduct synthesis, large sources of simple polysaccharides are required. Second generation feedstocks are a potential source of such material, as they are typically not food crops (such as many first generation bioproduct feedstocks) and therefore do not impact on human food security. Some plant species produce large amounts of dense lignocellulosic biomass and are able to grow in a wide range of conditions and environments (Hendriks and Zeeman 2009; Himmel et al. 2007). However, lignocellulosic biomass is highly recalcitrant (Himmel et al. 2007), requiring large investments of energy (with associated financial and waste costs) to access (Alvira et al. 2010).

The integration of Carbohydrate Active Enzymes (CAZymes) in the processing of lignocellulosic biomass is a promising strategy for reducing the difficulty and cost of biopolymer extraction (Mir et al. 2014; Turumtay 2015). CAZymes are active on oligosaccharides, polysaccharides and glycoconjugates. They consist of protein domains that are classified into a hierarchy of families based on their structure and function, including Glycoside Hydrolases (GHs), GlycosylTransferases (GTs), Carbohydrate-Binding Module (CBMs), Carbohydrate Esterases (CEs), Polysaccharide Lyases (PLs) and Auxiliary Activity families (AAs) (Cantarel et al. 2009; Lombard et al. 2014).

However, while CAZymes represent a useful tool for breaking down lignocellulosic biomass, they are generally not suited to the harsh conditions (especially extreme temperature) that form the basis of many industrial pretreatments of lignocellulosic biomass (Blumer-Schuette et al. 2014). As one approach to overcoming the problem of enzyme functional stability, thermostable CAZymes have been identified in and isolated from hyperthermophilic and extremely thermophilic organisms (i.e., organisms that grow optimally at temperatures exceeding 70°C (Leuschner and Antranikian 1995)). By mining genomes of extremely thermophilic organisms, it is possible to identify CAZymes that may operate effectively under the extreme conditions characteristic of industrial pre-treatments and other white biotechnological (i.e. industrial) applications. An added benefit of using extremely thermophilic CAZymes is that they may be expressed *in planta* with little fear of cytotoxicity, due to their very low activity at mesophilic temperatures (Mir et al. 2014). This approach can potentially reduce the need for exogenous enzyme loading, where accumulation of the expressed enzyme in the plant tissues

produces 'self-processing' plants, potentially increasing the efficiency of lignocellulose degradation at high temperatures (without impacting the normal growth and development of the plant at mesophilic temperatures (Mir et al. 2014)).

The field of synthetic biology has a long history, and through innovations such as the iGEM competition and Biobricks foundation (Smolke 2009; Vilanova and Porcar 2014) has established a registry of 'components' (including promoters, ribosome binding sites, protein coding sequences and terminators, among others; http://parts.igem.org) for the rational design and programming of systems in living cells (Endy 2005; Hartwell et al. 1999; Purnick and Weiss 2009). However, characterisation and incorporation of new individual components into the system remains an issue, and is a bottleneck to progress (Cameron et al. 2014). Most components are derived from wild-type systems, and are limited in function. For example, a xylanase which binds to cellulose may be desirable, in order to digest xylan closely associated with cellulose to make it more accessible to enzymic degradation, but xylanase CBMs do not typically bind well to cellulose (McCartney et al. 2006). Nevertheless, the concept of engineering rationally designed multi-component CAZymes, potentially capable of targeting lignocellulose biopolymer backbones and accessory structures would appear to offer considerable promise for lignocellulosic biomass processing. .

There have been a number of recent reviews summarising the industrial applications of extremely thermophilic CAZymes, either of specific enzyme classes (Atomi et al. 2011; Elleuche et al. 2015; Fatima and Hussain ; Nisha and Satyanarayana 2016), or with a focus on lignocellulose processing (Blumer-Schuette et al. 2014; Elleuche et al. 2014; Guerriero et al. 2015; Mir et al. 2014; Turumtay 2015; Urbieta et al. 2015). However, these reviews do not highlight the modular nature of hyperthermophilic CAZymes or how they might be used to tailor enzymes for lignocellulose deconstruction. In this review, we provide a survey of CAZyme modules with known lignocellulose degrading abilities derived from extremely thermophilic organisms, and explore how they might be applied in white biotechnology and enzyme engineering. We identify a list of predicted CAZyme domains (using publicly available proteomes from Ensemble Bacteria and HMMER protein domain prediction; Supplementary data 1) from the proteomes of a selection of extremely thermophilic organisms (Fig 1, Supplementary data 2). These domains cover a significant proportion of existing CAZyme families (Fig. 2) and comprise a wide range of activities and substrate specificities, as summarised in the CAZy database (www.cazy.org (Lombard et al. 2014)).
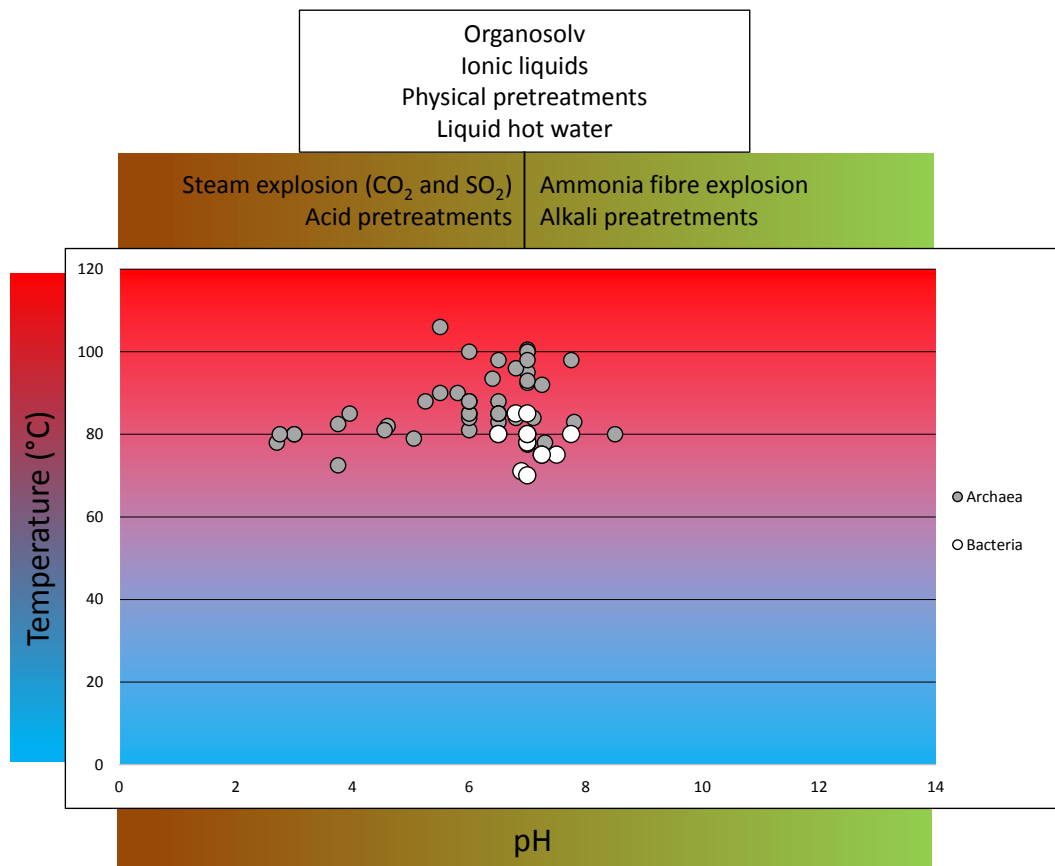
Figure 1 The spread of optimum growth temperature and pH of extremely thermophilic organisms covered in this review, as well as some of the main pretreatments (Alvira et al. 2010) associated with lignocellulose deconstruction.
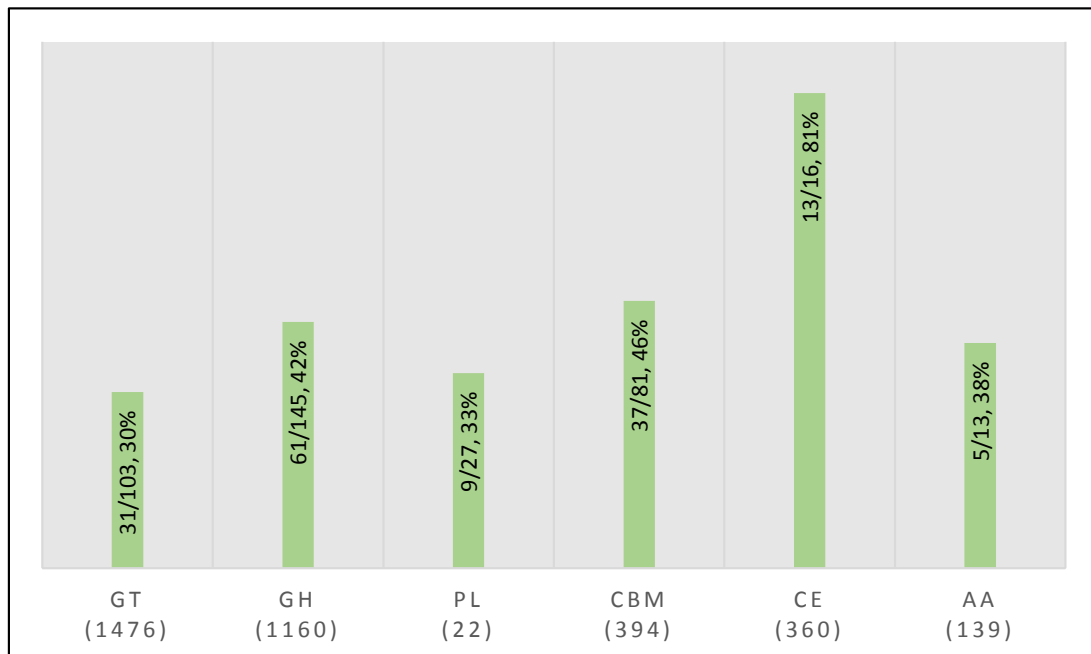

Figure 2 The overall coverage of CAZyme families identified from extremely thermophilic organisms, expressed as a percentage. The CAZyme class is listed on the X-axis. GT: Glycosyl Transferase, GH: Glycoside hydrolase, PL: Polysaccharide Lyase, CBM: Carbohydrate-Binding Module, CE: Carbohydrate Esterase and AA: Auxilliary Activity. The number listed below each class is the absolute count of domains identified in each class for extremely thermophilic organisms. The first number in the bars indicates the absolute proportion of all known CAZyme domain families present in the extremely thermophilic organisms and the second number is this proportion expressed as a percentage

CAZyme categories

CAZyme domains are divided into families, based on their structures (Cantarel et al. 2009; Lombard et al. 2014). This classification implies that two CAZyme domains that are members of the same family are likely to have similar activities and substrate specificities, but in reality this is not always the case. While broad trends are often seen within a CAZyme family, the diversity of activities, specificities and thermostabilities is high. This is evident in the GH3 domain family, where four GH3 enzymes from *Cellulomonas fimi* showed a range of activities on many xylo- and oligosaccharides (Gao and Wakarchuk 2014) even though no other catalytic domains were apparent. The GH5 domain family also displays considerable functional diversity, with recent studies highlighting specificities for substrates such as cellulose (Huy et al. 2016; Valadares et al. 2016; Wang et al. 2016b), xylans and xyloglucans (dos Santos et al. 2015; Ghatge et al. 2014) mannans (Tóth et al. 2016; Zang et al. 2015) and glycoceramides (Han et al. 2017b). Occasionally, new CAZyme domain families or rare variants of existing families are also discovered, such as a xylan degrading (Corrêa et al. 2012) or multifunctional (Morrison et al. 2016) GH39 enzyme, or the recently defined GH116 family, which was shown to act on glucosylceramide, N-acetylglucosaminides and xylosides (Cobucci-Ponzano et al. 2010; Ferrara et al. 2014). If each CAZyme domain is viewed as a potential building block for the rational design and synthesis of enzymes for a range of applications, then identification and characterisation of new CAZyme domains from both known and novel families will further supplement our toolset for enzyme design.

'Omics' technologies are excellent methods for expanding the CAZyme repertoire. By investigating the (meta)genomes, (meta)transcriptomes, (meta)proteomes and (meta)secretomes of organisms and communities which process lignocellulose (Kuuskeri et al. 2016; López-Mondéjar et al. 2016; Montella et al. 2017; Schneider et al. 2016; Solomon et al. 2016; Wang et al. 2016a), a full complement of lignocellulose degrading CAZymes may be identified. Additionally, CAZyme genes from organisms which synthesise lignocellulose (such as plants (Geisler-Lee et al. 2006; Pinard et al. 2015) and some bacteria (reference)) could allow for diversified modification of biopolymers (such as adding side chains or chemical groups, and altering the structure of xylan (Abramson et al. 2010)).

Cataloguing each domain individually could provide a comprehensive toolbox for enzyme design, but the capability of such a toolbox could be drastically increased by determining the underlying mechanisms of domain variety and implementing them in protein design. This would allow for fine-tuning of synthetic enzymes to specific processes. A key technique for investigating differences in mechanisms of action is protein

crystallography, which uses high quality structural data to identify mechanisms of binding or stability between and within CAZyme domains (Czjzek and Ficko-Blean 2017). Relatively few CAZyme domains have been structurally resolved at high resolution (www.cazy.org), but every CAZyme class is based on at least one resolved structure.

Recent examples of mechanistic insights from structural analyses include new variants of rare domain structures (Godoy et al. 2016), general mechanisms of substrate binding (as seen in studies on CBM35 xylanases (Sainz-Polo et al. 2014a; Sainz-Polo et al. 2014b; Valenzuela et al. 2012) and GH30 xylanases (Sainz-Polo et al. 2014a; Sainz-Polo et al. 2014b; Verma and Goyal 2014; Verma et al. 2013)) as well as interactions with specific substrates (Attia et al. 2016; dos Santos et al. 2015; Pires et al. 2017).

The value of structural data is not limited to understanding interactions of domains with substrates. In some cases it can help to elucidate how enzymes behave under certain conditions (such as the mechanistic basis for glucose tolerant and intolerant GH1 domains (Yang et al. 2015)). Structural data can reveal interactions between domains in a single enzyme. The resolved structure of Xyn10C from *Paenibacillus barcinonensis* (Sainz-Polo et al. 2015a) is the first structure of an enzyme with two tandem CBMs, showing the architecture and interaction of multiple domains in an enzyme.

Engineering CAZyme domains

Studying CAZyme domains and the mechanisms by which they perform their functions is the first step towards informed rational CAZyme design. While it is possible for a protein to have multiple domains with different substrate targets (Sainz-Polo et al. 2015b), enzyme design constitutes more than simply choosing domains with desirable functions and combining them in an arbitrary manner (André et al. 2014; Elleuche 2015). For example, the addition of a CBM2 , CBM3 or CBM4 CAZyme domains to enzymes increased binding affinity, thermostability and enzyme activity ((Voutilainen et al. 2014; Duan et al. 2017; Foumani et al. 2015; Liu and Ding 2016), . whereas the removal of a CBM domain (had a deleterious effect on enzyme performance, with reduced activity and thermostability(Ghatge et al. 2014; Li et al. 2015; Yi et al. 2013).The relationship between domain functions is not necessarily additive. Whole CAZymes can behave in a synergistic manner (Chung et al. 2015; Liu and Ding 2016) and CAZyme domains have also been shown to functionally synergise (Diogo et al. 2015; Liu and Ding 2016). However, the effects of domain addition or deletion are not necessarily predictable. For example, separating two domains that co-occur in a wild-type enzyme (such as the CBM46 domain from

*Bacillus halodurans Bh*Cel5b) can reduce or abolish catalytic activity of both (Venditto et al. 2015). Conversely, some domains operate more efficiently when separated from each other (Khan et al. 2013). The addition of a known thermostabilizing domain such as CBM22 (Khan et al. 2013; Lee et al. 1993) can reduce the thermostability of the protein construct while increasing catalytic efficiency (Araki et al. 2006). These examples emphasise the complex interactions that occur between CAZyme domains within an enzyme.

Table 1 Quantities of CAZyme families identified by this review and the main substrates they are known to target

CAZyme family[a]	Number[b]	Main substrates[c]
CBM2	3	Cellulose, chitin and xylan.
CBM3	18	Cellulose and chitin
CBM4	18	Xylan, β-1,3-glucan, β-1,3-1,4-glucan, β-1,6-glucan and amorphous cellulose
CBM6	1	Amorphous cellulose, xylan, β-1,3-glucan, β-1,3-1,4-glucan, and β-1,4-glucan
CBM9	19	Xylan and cellulose
CBM13	3	Mannose and xylan
CBM16	3	Cellulose and glucomannan
CBM22	37	Xylan and mixed β-1,3/β-1,4-glucans
CBM28	2	Non-crystalline cellulose, cellooligosaccharides, and β-(1,3)(1,4)-glucans
CBM35	15	Xylan, decorated soluble mannans, mannooligosaccharides and β-galactan.
CBM36	1	Xylans and xylooligosaccharides
CBM37	8	Xylan, chitin, microcrystalline and phosphoric-acid swollen cellulose, alfalfa cell walls, banana stem and wheat straw
CBM44	28	Cellulose and xyloglucan
CBM46	1	Cellulose
CBM54	9	Xylan, yeast cell wall glucan and chitin
CBM60	1	Xylan
CBM63	1	Cellulose
GH10	24	Cellulose and xylan
GH11	1	Xylan
GH1	54	Cellulose, xylan, mannan and xyloglucan
GH3	30	Cellulose and xylan
GH5	31	Cellulose, xylan, mannan, lichenin, chitosan, xyloglucan and arabinoxylan
GH8	1	Chitosan, cellulose, lichenin and xylan
GH9	3	Cellulose, lichenin, xyloglucan
GH12	28	Cellulose, xylan and xyloglucan
GH16	10	Cellulose, xylan, xyloglucan, lichenin and chitin
GH26	5	Mannan, xylan and cellulose

[a]The CAZyme family designation
[b]The number of domains identified from extremely thermophilic proteins using HMMER
[c]The main substrates on which these domains act

The most obvious and immediate use for a toolbox of CAZyme domains would be to provide a catalogue of parts from which custom enzymes may be assembled. However, as the toolbox expands, it may begin to fill a substantially more significant function - providing a set of protein scaffolds for rational engineering on new functionalities. To date, reports of rationally engineered CAZymes are limited, but two main strategies are being employed: directed evolution (DE) and rational design (Davids et al. 2013).

Using DE, thermostability of GH10 Xyn III from *Trichoderma reesei* was enhanced (Matsuzawa et al. 2016). Similarly, a GH51 α-l-arabinofuranosidase has been engineered for higher transglycosylating activity (Arab-Jaziri et al. 2013, 2015) and reduced secondary hydrolysis of transglycosylation products (Bissaro et al. 2014). While DE is an attractive option for protein engineering, widespread use of the technique is impeded by difficulties with screening and selection of variants with desirable traits (Turner 2009).

To overcome this problem, a number of strategies have been developed. By limiting analysis to mutations in specific residues (such as those in the active sites of proteins), it is possible to reduce the number of variants to be screened. CASTing, or Combinatorial Active Site Testing is one method by which this can be achieved (Reetz et al. 2006). This approach require knowledge of the crystal structure of the protein to accurately identify the sites for modification. If the crystal structure is not available, high throughput screens are used. Techniques such as ribosome display (Gan and Jewett 2016), mRNA display (Horiya et al. 2017), Yeast display (Traxlmayr and Shusta 2017) and Phage-based techniques (Brödel et al. 2017) may be used in combination with in vitro compartmentalization-based fluorescence-activated cell sorting (IVC-FACS) to identify and pool variants with desirable properties (Ma et al. 2016). Finally, colorimetric assays based on enzyme reaction mechanisms can be adapted to high throughput applications (Smart et al. 2017). Recently, a kit was developed that contains chromogenic substrates that can be used to test the activities of carbohydrate degrading enzymes that can be multiplexed in a 96 well format (Schückel et al. 2016). Although substantial progress has been made, selection and screening of DE clone libraries remains a challenge and continues to be an area of active research and innovation (Klenk et al. 2016; Lin et al. 2017; Ma et al. 2016; Reetz 2017).

The limitations of DE (Turner 2009) may be overcome through rational design (directly targeting and mutating specific residues of a protein to obtain a desired effect), although this approach requires in-depth knowledge of the structure and mechanisms of the protein. Using this method, various CAZyme properties have been altered; e.g., increasing both the thermal stability and optimal catalytic temperature of GH10 (de Souza et al. 2016) and GH11 (Han et al. 2017a) xylanases by the introduction of a mutations (by site-directed mutagenesis) to influence disulphide bond formation, salt bridges and the ratio of acidic to basic amino acids (de Souza et al. 2016; Han et al. 2017a).

Enzyme engineering of CAZymes to confer new functional characteristics has been applied with some success. Altering active site architecture through mutation (e.g. W22Y) changed the binding properties of *T. reesei* GH12 enzyme *Tr*Cel12A, resulting in expanded substrate specificity (Zhang et al. 2015). Similarly, a GH1 β-glucosidase from *Thermus thermophilus* was converted to a *trans*-β-acetylglucosaminidase by mutation of the N163 and E338 residues in the active site to remove steric conflicts with the *N*-acetyl-d-glucosamine substrate (André-Miral et al. 2015), This is an especially significant result, considering that no native GH1 family protein exhibits *trans*-β-acetylglucosaminidase activity. Additionally, a single predicted amino acid change in a *Geobacillus stearothermophilus* β-xylosidase (Y509E) was enough to confer new exo-xylanase functionality to the enzyme (Huang et al. 2014).

Applying the toolbox: CAZyme domains for *in planta* lignocellulose degradation

The efficient breakdown of lignocellulose can be achieved using extremes of temperature, pressure and pH, which differ depending on which biopolymer is the target of extraction (Alvira et al. 2010). While supplementing physical and chemical pre-treatment processes with enzymes can potentially reduce the economic and energy investment required for biopolymer extraction (Blumer-Schuette et al. 2014), in planta expression of these enzymes (as opposed to external enzyme loading) may also be beneficial (Mir et al. 2014, 2017). Expressing thermostable enzymes directly in the plant tissue should not disrupt normal growth and development of the biomass, due to inactivity of the enzymes at lower temperature (Mir et al. 2014; Montalvo-Rodriguez et al. 2000). On heating, the harvested biomass undergoes some autohydrolysis (Bhatia et al. 2017; Mir et al. 2014). This strategy has been shown to be effective in first generation feedstocks (Kim et al. 2016).

However, studies on the heterologous in planta expression of thermostable CAZymes have shown that while this strategy does improve digestibility, localisation of the product is important (Castiglia et al. 2016; Kim et al. 2016). Expressing a protein in the wrong cellular compartment can lead to deleterious effects. For example, plastid targeted expression of a thermostable endoglucanase in tobacco resulted in binding of the recombinant gene product to thylakoid membranes and a negative impact on plastid development (Castiglia et al. 2016). Conversely, enzymes may perform better, or accumulate more if expressed in the correct locale [such as Xyl10b from *Thermotoga maritima MSB8*, which had higher yield and specific activity when targeted to the apoplast vs the chloroplast (Kim et al. 2016)]. Fusion to a CBM is one strategy through which translocation of a protein product to a desired location may be achieved (Oliveira et al. 2015).

Conclusions

Hyperthermophilic and extremely thermophilic organisms are a potentially valuable source of thermostable CAZyme domains for industrial use (Blumer-Schuette et al. 2014). The ability to degrade lignocellulosic substrates is surprisingly common in hyperthermophilic and extremely thermophilic organisms, despite the oligotrophic nature the ecological niches these organisms typically inhabit (Chaban et al. 2006; Rothschild and Mancinelli 2001). There are now numerous reports of highly stable CAZymes, across most of the CAZyme families and we have provided a list of putative thermostable CAZyme domains that can break down cellulose and xylan (Table 2, Fig. 3), as well as other important lignocellulosic biopolymers (Online Resource 3).

However, many issues relating to the function and engineering of these enzymes remain unresolved. Modifications to CAZymes can have unpredictable consequences and the achievable limits for engineered catalytic ability are not known. Additionally, considering the functional diversity in some CAZyme families, it is unclear whether molecular modelling and docking studies on some enzymes in a family may be applied to others. Finally, while producing self-processing lignocellulosic biomass is an attractive prospect for industry, the impact of *in planta* production of enzymes on other performance factors (such as disease resistance) of plant growth need to be determined.

CAZyme domains from extremely thermophilic organisms provide a toolbox which can be used to design enzymes suited to a range of industrial processes. As our understanding of protein domains and enzyme engineering increases, so will the value of such a toolbox. These domains provide a way to modify existing enzymes and are also highly thermostable scaffolds for further modifications. Studying these domains and their

interactions within proteins could eventually facilitate de novo design and synthesis of new highly thermostable

and highly catalytic proteins.

Table 2. Quantities of CAZyme families identified by this review and the main substrates they are known to target, according to the CAZy database (http://www.cazy.org; Lombard et al. 2014)

CAZyme family[a]	Number[b]	Main substrates[c]
CBM2	3	Cellulose, chitin and xylan.
CBM3	18	Cellulose and chitin
CBM4	18	Xylan, β-1,3-glucan, β-1,3-1,4-glucan, β-1,6-glucan and amorphous cellulose
CBM6	1	Amorphous cellulose, xylan, β-1,3-glucan, β-1,3-1,4-glucan, and β-1,4-glucan
CBM9	19	Xylan and cellulose
CBM13	3	Mannose and xylan
CBM16	3	Cellulose and glucomannan
CBM22	37	Xylan and mixed β-1,3/β-1,4-glucans
CBM28	2	Non-crystalline cellulose, cellooligosaccharides, and β-(1,3)(1,4)-glucans
CBM35	15	Xylan, decorated soluble mannans, mannooligosaccharides and β-galactan.
CBM36	1	Xylans and xylooligosaccharides
CBM37	8	Xylan, chitin, microcrystalline and phosphoric-acid swollen cellulose, alfalfa cell walls, banana stem and wheat straw
CBM44	28	Cellulose and xyloglucan
CBM46	1	Cellulose
CBM54	9	Xylan, yeast cell wall glucan and chitin
CBM60	1	Xylan
CBM63	1	Cellulose
GH10	24	Cellulose and xylan
GH11	1	Xylan
GH1	54	Cellulose, xylan, mannan and xyloglucan
GH3	30	Cellulose and xylan
GH5	31	Cellulose, xylan, mannan, lichenin, chitosan, xyloglucan and arabinoxylan
GH8	1	Chitosan, cellulose, lichenin and xylan
GH9	3	Cellulose, lichenin, xyloglucan
GH12	28	Cellulose, xylan and xyloglucan
GH16	10	Cellulose, xylan, xyloglucan, lichenin and chitin
GH26	5	Mannan, xylan and cellulose

[a]The CAZyme family designation

[b]The number of domains identified from extremely thermophilic proteins using HMMER

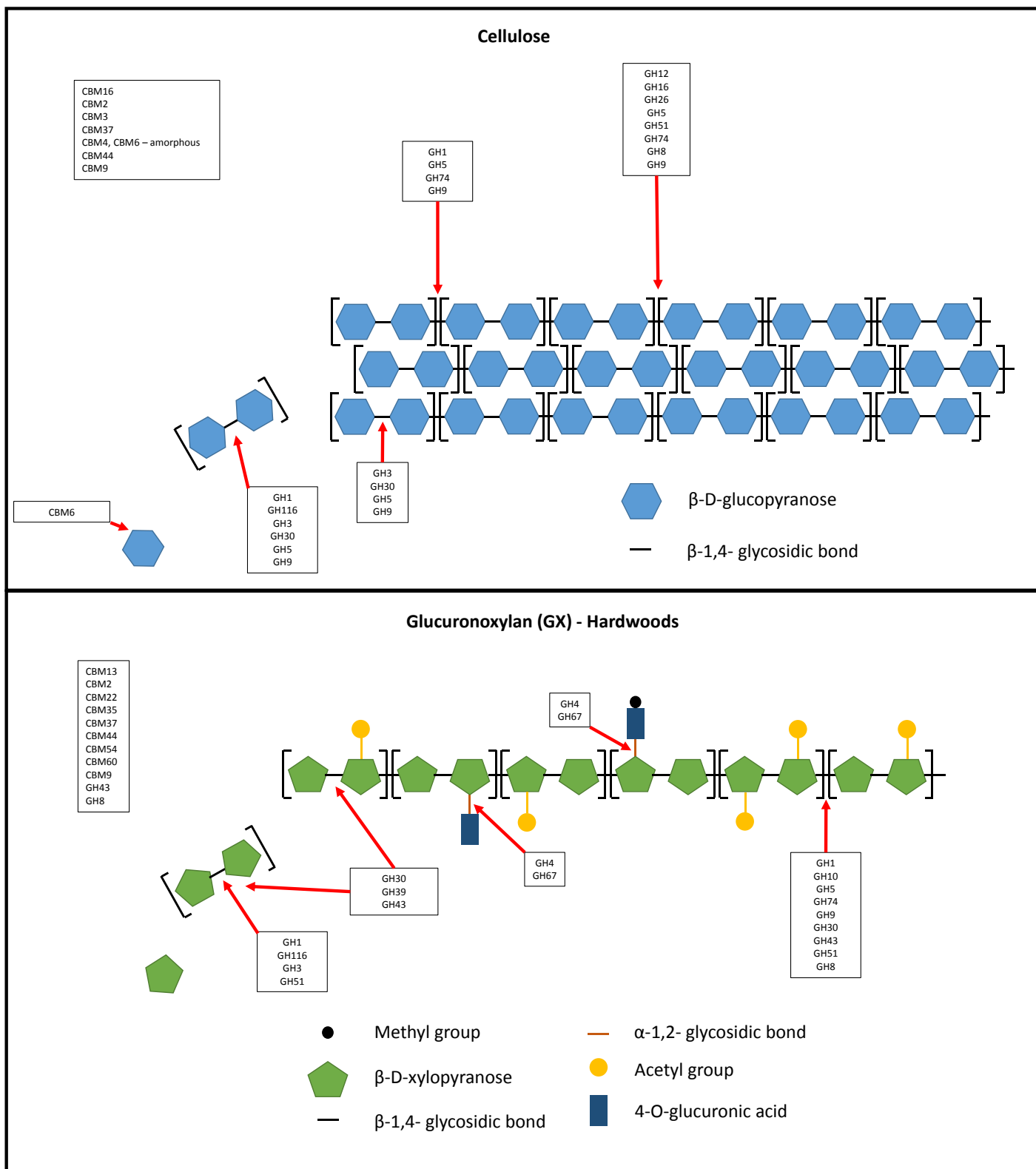[c]The main substrates on which these domains act

11

Figure 3 Schematic representation of important plant secondary cell wall biopolymers, as well as the CAZyme domains which can degrade them covered by this review. The name of the biopolymer is listed at the top of the figure. The box in the top left of the diagram indicates CAZyme domain families which have activities on the biopolymer, but have no record of specific interactions. The red arrows indicate specific areas of activity. The key is located at the bottom of each diagram

Acknowledgements

References

Abramson M, Shoseyov O, Shani Z (2010) Plant cell wall reconstruction toward improved lignocellulosic production and processability Plant Science 178:61-72

Alvira P, Tomás-Pejó E, Ballesteros M, Negro M (2010) Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review Bioresource technology 101:4851-4861

André-Miral C, Koné FM, Solleux C, Grandjean C, Dion M, Tran V, Tellier C (2015) De novo design of a trans-β-N-acetylglucosaminidase activity from a GH1 β-glycosidase by mechanism engineering Glycobiology 25:394-402

André I, Potocki-Véronèse G, Barbe S, Moulis C, Remaud-Siméon M (2014) CAZyme discovery and design for sweet dreams Current Opinion in Chemical Biology 19:17-24

Arab-Jaziri F et al. (2013) Engineering transglycosidase activity into a GH51 α-l-arabinofuranosidase New biotechnology 30:536-544

Arab-Jaziri F, Bissaro B, Tellier C, Dion M, Fauré R, O'Donohue MJ (2015) Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α-l-arabinofuranosidase Carbohydrate Research 401:64-72

Araki R, Karita S, Tanaka A, Kimura T, Sakka K (2006) Effect of family 22 carbohydrate-binding module on the thermostability of Xyn10B catalytic module from Clostridium stercorarium Bioscience Biotechnology and Biochemistry 70:3039

Atomi H, Sato T, Kanai T (2011) Application of hyperthermophiles and their enzymes Current Opinion in Biotechnology 22:618-626 doi:http://dx.doi.org/10.1016/j.copbio.2011.06.010

Attia M, Stepper J, Davies GJ, Brumer H (2016) Functional and structural characterization of a potent GH74 endo-xyloglucanase from the soil saprophyte Cellvibrio japonicus unravels the first step of xyloglucan degradation FEBS Journal 283:1701-1719

Bhatia R, Gallagher JA, Gomez LD, Bosch M (2017) Genetic engineering of grass cell wall polysaccharides for biorefining Plant Biotechnology Journal

Bissaro B et al. (2014) Mutation of a pH-modulating residue in a GH51 α-l-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products Biochimica et Biophysica Acta (BBA)-General Subjects 1840:626-636

Blumer-Schuette SE et al. (2014) Thermophilic lignocellulose deconstruction FEMS microbiology reviews 38:393-448

Brock TD (2012) Thermophilic microorganisms and life at high temperatures. Springer Science & Business Media,

Cameron DE, Bashor CJ, Collins JJ (2014) A brief history of synthetic biology Nature Reviews Microbiology 12:381-390

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics Nucleic acids research 37:D233-D238

Castiglia D et al. (2016) High-level expression of thermostable cellulolytic enzymes in tobacco transplastomic plants and their use in hydrolysis of an industrially pretreated Arundo donax L. biomass Biotechnology for Biofuels 9:1

Chaban B, Ng SY, Jarrell KF (2006) Archaeal habitats-from the extreme to the ordinary Canadian journal of microbiology 52:73-116

Chung D, Young J, Cha M, Brunecky R, Bomble YJ, Himmel ME, Westpheling J (2015) Expression of the Acidothermus cellulolyticus E1 endoglucanase in Caldicellulosiruptor bescii enhances its ability to deconstruct crystalline cellulose Biotechnology for biofuels 8:1

Cobucci-Ponzano B et al. (2010) A New Archaeal β-Glycosidase from Sulfolobus solfataricus SEEDING A NOVEL RETAINING β-GLYCAN-SPECIFIC GLYCOSIDE HYDROLASE FAMILY ALONG WITH THE HUMAN NON-LYSOSOMAL GLUCOSYLCERAMIDASE GBA2 Journal of Biological Chemistry 285:20691-20703

Corrêa JM et al. (2012) Expression and characterization of a GH39 β-xylosidase II from Caulobacter crescentus Applied Biochemistry and biotechnology 168:2218-2229

Couturier M, Feliu J, Bozonnet S, Roussel A, Berrin J-G (2013) Molecular engineering of fungal GH5 and GH26 beta-(1, 4)-mannanases toward improvement of enzyme activity PloS one 8:e79800

Czjzek M, Ficko-Blean E (2017) Probing the Complex Architecture of Multimodular Carbohydrate-Active Enzymes Using a Combination of Small Angle X-Ray Scattering and X-Ray Crystallography Protein-Carbohydrate Interactions: Methods and Protocols:239-253

Davids T, Schmidt M, Böttcher D, Bornscheuer UT (2013) Strategies for the discovery and engineering of enzymes for biocatalysis Current opinion in chemical biology 17:215-220

de Souza AR et al. (2016) Engineering increased thermostability in the GH-10 endo-1, 4-β-xylanase from Thermoascus aurantiacus CBMAI 756 International Journal of Biological Macromolecules 93:20-26

Diogo JA et al. (2015) Development of a chimeric hemicellulase to enhance the xylose production and thermotolerance Enzyme and Microbial technology 69:31-37

dos Santos CR, Cordeiro RL, Wong DW, Murakami MT (2015) Structural Basis for Xyloglucan Specificity and α-d-Xyl p (1→6)-d-Glc p Recognition at the− 1 Subsite within the GH5 Family Biochemistry 54:1930-1942

Duan C-J, Huang M-Y, Pang H, Zhao J, Wu C-X, Feng J-X (2017) Characterization of a novel theme C glycoside hydrolase family 9 cellulase and its CBM-chimeric enzymes Applied Microbiology and Biotechnology:1-15

Elleuche S (2015) Bringing functions together with fusion enzymes—from nature's inventions to biotechnological applications Applied microbiology and biotechnology 99:1545-1556

Elleuche S, Schäfers C, Blank S, Schröder C, Antranikian G (2015) Exploration of extremophiles for high temperature biotechnological processes Current opinion in microbiology 25:113-119

Elleuche S, Schröder C, Sahm K, Antranikian G (2014) Extremozymes—biocatalysts with unique properties from extremophilic microorganisms Current Opinion in Biotechnology 29:116-123

Endy D (2005) Foundations for engineering biology Nature 438:449-453

Espina G, Eley K, Pompidor G, Schneider TR, Crennell SJ, Danson MJ (2014) A novel β-xylosidase structure from Geobacillus thermoglucosidasius: the first crystal structure of a glycoside hydrolase family GH52 enzyme reveals unpredicted similarity to other glycoside hydrolase folds Acta Crystallographica Section D: Biological Crystallography 70:1366-1374

Fatima B, Hussain Z XYLOSE ISOMERASES FROM THERMOTOGALES

Ferrara MC, Cobucci-Ponzano B, Carpentieri A, Henrissat B, Rossi M, Amoresano A, Moracci M (2014) The identification and molecular characterization of the first archaeal bifunctional exo-β-glucosidase/N-acetyl-β-glucosaminidase demonstrate that family GH116 is made of three functionally distinct subfamilies Biochimica et Biophysica Acta (BBA)-General Subjects 1840:367-377

Foumani M, Vuong TV, MacCormick B, Master ER (2015) Enhanced Polysaccharide Binding and Activity on Linear β-Glucans through Addition of Carbohydrate-Binding Modules to Either Terminus of a Glucooligosaccharide Oxidase PloS one 10:e0125398

Gao J, Wakarchuk W (2014) Characterization of five β-glycoside hydrolases from Cellulomonas fimi ATCC 484 Journal of bacteriology 196:4103-4110

Geisler-Lee J et al. (2006) Poplar carbohydrate-active enzymes. Gene identification and expression analyses Plant Physiology 140:946-962

Ghatge SS et al. (2014) Characterization of modular bifunctional processive endoglucanase Cel5 from Hahella chejuensis KCTC 2396 Applied Microbiology and Biotechnology 98:4421-4435

Godoy AS, de Lima MZ, Camilo CM, Polikarpov I (2016) Crystal structure of a putative exo-β-1, 3-galactanase from Bifidobacterium bifidum S17 Acta Crystallographica Section F: Structural Biology Communications 72:288-293

Guerriero G, Hausman J-F, Strauss J, Ertan H, Siddiqui KS (2015) Destructuring plant biomass: Focus on fungal and extremophilic cell wall hydrolases Plant Science 234:180-193

Han N, Miao H, Ding J, Li J, Mu Y, Zhou J, Huang Z (2017a) Improving the thermostability of a fungal GH11 xylanase via site-directed mutagenesis guided by sequence and structural analysis Biotechnology for Biofuels 10:133

Han Y-B, Chen L-Q, Li Z, Tan Y-M, Feng Y, Yang G-Y (2017b) Structural Insights into the Broad Substrate Specificity of a Novel Endoglycoceramidase I Belonging to a New Subfamily of GH5 Glycosidases Journal of Biological Chemistry 292:4789-4800

Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology Nature 402:C47-C52

Hendriks A, Zeeman G (2009) Pretreatments to enhance the digestibility of lignocellulosic biomass Bioresource technology 100:10-18

Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production science 315:804-807

Huang Z, Liu X, Zhang S, Liu Z (2014) GH52 xylosidase from Geobacillus stearothermophilus: characterization and introduction of xylanase activity by site-directed mutagenesis of Tyr509 Journal of industrial microbiology & biotechnology 41:65-74

Huy ND et al. (2016) Characterization of a novel manganese dependent endoglucanase belongs in GH family 5 from Phanerochaete chrysosporium Journal of bioscience and bioengineering 121:154-159

Khan MIM, Sajjad M, Sadaf S, Zafar R, Niazi UH, Akhtar MW (2013) The nature of the carbohydrate binding module determines the catalytic efficiency of xylanase Z of Clostridium thermocellum Journal of Biotechnology 168:403-408

Khatri V, Hébert-Ouellet Y, Meddeb-Mouelhi F, Beauregard M (2016) Specific tracking of xylan using fluorescent-tagged carbohydrate-binding module 15 as molecular probe Biotechnology for biofuels 9:1

Kim JY, Nong G, Rice JD, Gallo M, Preston JF, Altpeter F (2016) In planta production and characterization of a hyperthermostable GH10 xylanase in transgenic sugarcane Plant Molecular Biology:1-14

Kuuskeri J et al. (2016) Time-scale dynamics of proteome and transcriptome of the white-rot fungus Phlebia radiata: growth on spruce wood and decay effect on lignocellulose Biotechnology for Biofuels 9:192

Lafond M, Sulzenbacher G, Freyd T, Henrissat B, Berrin J-G, Garron M-L (2016) The quaternary structure of a glycoside hydrolase dictates specificity toward β-glucans Journal of Biological Chemistry 291:7183-7194

Lee Y-E, Lowe S, Henrissat B, Zeikus JG (1993) Characterization of the active site and thermostability regions of endoxylanase from Thermoanaerobacterium saccharolyticum B6A-RI Journal of bacteriology 175:5890-5898

Leuschner C, Antranikian G (1995) Heat-stable enzymes from extremely thermophilic and hyperthermophilic microorganisms World Journal of Microbiology and Biotechnology 11:95-114

Li S, Yang X, Bao M, Wu Y, Yu W, Han F (2015) Family 13 carbohydrate-binding module of alginate lyase from Agarivorans sp. L11 enhances its catalytic efficiency and thermostability, and alters its substrate preference and product distribution FEMS microbiology letters 362:fnv054

Liu S, Ding S (2016) Replacement of carbohydrate binding modules improves acetyl xylan esterase activity and its synergistic hydrolysis of different substrates with xylanase BMC biotechnology 16:73

Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013 Nucleic acids research 42:D490-D495

López-Mondéjar R, Zühlke D, Větrovský T, Becher D, Riedel K, Baldrian P (2016) Decoding the complete arsenal for cellulose and hemicellulose deconstruction in the highly efficient cellulose decomposer Paenibacillus O199 Biotechnology for biofuels 9:104

Matsuzawa T, Kaneko S, Yaoi K (2016) Improvement of thermostability and activity of Trichoderma reesei endo-xylanase Xyn III on insoluble substrates Applied microbiology and biotechnology 100:8043-8051

McCartney L, Blake AW, Flint J, Bolam DN, Boraston AB, Gilbert HJ, Knox JP (2006) Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules Proceedings of the National Academy of Sciences of the United States of America 103:4765-4770

McKee LS et al. (2012) Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains Proceedings of the National Academy of Sciences 109:6537-6542

Mir BA, Mewalal R, Mizrachi E, Myburg AA, Cowan DA (2014) Recombinant hyperthermophilic enzyme expression in plants: a novel approach for lignocellulose digestion Trends in biotechnology 32:281-289

Montella S, Ventorino V, Lombard V, Henrissat B, Pepe O, Faraco V (2017) Discovery of genes coding for carbohydrate-active enzyme by metagenomic analysis of lignocellulosic biomasses Scientific Reports 7

Morrison JM, Elshahed MS, Youssef N (2016) A multifunctional GH39 glycoside hydrolase from the anaerobic gut fungus Orpinomyces sp. strain C1A PeerJ 4:e2289

Naik SN, Goud VV, Rout PK, Dalai AK (2010) Production of first and second generation biofuels: A comprehensive review Renewable and Sustainable Energy Reviews 14:578-597 doi:http://dx.doi.org/10.1016/j.rser.2009.10.003

Nisha M, Satyanarayana T (2016) Characteristics, protein engineering and applications of microbial thermostable pullulanases and pullulan hydrolases Applied microbiology and biotechnology:1-19

Oliveira C, Carvalho V, Domingues L, Gama FM (2015) Recombinant CBM-fusion technology—applications overview Biotechnology advances 33:358-369

Pinard D et al. (2015) Comparative analysis of plant carbohydrate active enZymes and their role in xylogenesis Bmc Genomics 16:402

Pires VM et al. (2017) Stability and ligand promiscuity of type A carbohydrate-binding modules are illustrated by the structure of Spirochaeta thermophila StCBM64C Journal of Biological Chemistry:jbc. M116. 767541

Purnick PE, Weiss R (2009) The second wave of synthetic biology: from modules to systems Nature reviews Molecular cell biology 10:410-422

Rothschild LJ, Mancinelli RL (2001) Life in extreme environments Nature 409:1092-1101

Sainz-Polo MA, González B, Menéndez M, Pastor FJ, Sanz-Aparicio J (2015a) Exploring Multimodularity in Plant Cell Wall Deconstruction STRUCTURAL AND FUNCTIONAL ANALYSIS OF Xyn10C

CONTAINING THE CBM22-1–CBM22-2 TANDEM Journal of Biological Chemistry 290:17116-17130

Sainz-Polo MA, González B, Menéndez M, Pastor FJ, Sanz-Aparicio J (2015b) Exploring multimodularity in plant cell wall deconstruction: structural and functional analysis of Xyn10C containing the CBM22-1-CBM22-2 tandem Journal of Biological Chemistry:jbc. M115. 659300

Sainz-Polo MA, Valenzuela SV, González B, Pastor FJ, Sanz-Aparicio J (2014a) Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity Journal of Biological Chemistry 289:31088-31101

Sainz-Polo MÁ, Valenzuela SV, Pastor FJ, Sanz-Aparicio J (2014b) Crystallization and preliminary X-ray diffraction analysis of Xyn30D from Paenibacillus barcinonensis Acta Crystallographica Section F: Structural Biology Communications 70:963-966

Schneider WDH et al. (2016) Penicillium echinulatum secretome analysis reveals the fungi potential for degradation of lignocellulosic biomass Biotechnology for biofuels 9:66

Smolke CD (2009) Building outside of the box: iGEM and the BioBricks Foundation Nature biotechnology 27:1099-1102

Solomon KV et al. (2016) Early-branching gut fungi possess a large, comprehensive array of biomass-degrading enzymes Science 351:1192-1195

Tóth Á et al. (2016) Cloning, Expression and Biochemical Characterization of Endomannanases from Thermobifida Species Isolated from Different Niches PloS one 11:e0155769

Turner NJ (2009) Directed evolution drives the next generation of biocatalysts Nature chemical biology 5:567-573

Turumtay H (2015) Cell Wall Engineering by Heterologous Expression of Cell Wall-Degrading Enzymes for Better Conversion of Lignocellulosic Biomass into Biofuels BioEnergy Research:1-15

Urbieta MS, Donati ER, Chan K-G, Shahar S, Sin LL, Goh KM (2015) Thermophiles in the genomic era: biodiversity, science, and applications Biotechnology advances 33:633-647

Valadares F et al. (2016) Exploring glycoside hydrolases and accessory proteins from wood decay fungi to enhance sugarcane bagasse saccharification Biotechnology for biofuels 9:110

Valenzuela SV, Diaz P, Pastor FJ (2012) Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module Applied and environmental microbiology 78:3923-3931

Venditto I et al. (2015) Family 46 Carbohydrate-binding Modules Contribute to the Enzymatic Hydrolysis of Xyloglucan and β-1, 3–1, 4-Glucans through Distinct Mechanisms Journal of Biological Chemistry 290:10572-10586

Verma AK, Goyal A (2014) In silico structural characterization and molecular docking studies of first glucuronoxylan-xylanohydrolase (Xyn30A) from family 30 glycosyl hydrolase (GH30) from Clostridium thermocellum Molecular biology 48:278-286

Verma AK et al. (2013) Overexpression, crystallization and preliminary X-ray crystallographic analysis of glucuronoxylan xylanohydrolase (Xyn30A) from Clostridium thermocellum Acta Crystallographica Section F: Structural Biology and Crystallization Communications 69:1440-1442

Vilanova C, Porcar M (2014) iGEM 2.0 [mdash] refoundations for engineering biology Nature biotechnology 32:420-424

Voutilainen SP, Nurmi-Rantala S, Penttilä M, Koivula A (2014) Engineering chimeric thermostable GH7 cellobiohydrolases in Saccharomyces cerevisiae Applied Microbiology and Biotechnology 98:2991-3001

Wang C, Dong D, Wang H, Müller K, Qin Y, Wang H, Wu W (2016a) Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition Biotechnology for biofuels 9:22

Wang Y, Yu W, Han F (2016b) Expression and characterization of a cold-adapted, thermotolerant and denaturant-stable GH5 endoglucanase Celal_2753 that withstands boiling from the psychrophilic bacterium Cellulophaga algicola IC166T Biotechnology letters 38:285-290

Yang Y et al. (2015) A mechanism of glucose tolerance and stimulation of GH1 β-glucosidases Scientific reports 5

Yi Z, Su X, Revindran V, Mackie RI, Cann I (2013) Molecular and biochemical analyses of CbCel9A/Cel48A, a highly secreted multi-modular cellulase by Caldicellulosiruptor bescii during growth on crystalline cellulose PloS one 8:e84172

Zang H et al. (2015) A novel thermostable GH5_7 β-mannanase from Bacillus pumilus GBSW19 and its application in manno-oligosaccharides (MOS) production Enzyme and Microbial technology 78:1-9

Zhang X et al. (2015) Subsite-specific contributions of different aromatic residues in the active site architecture of glycoside hydrolase family 12 Scientific reports 5