View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repository of the Academy's Library

Acta Alimentaria, Vol. 46 (3), pp. 323–335 (2017)
DOI: 10.1556/066.2016.0014

PLANT POLYSACCHARIDE DEGRADING ENZYME SYSTEM OF *THERMOBIFIDA CELLULOSILYTICA* TB100^T REVEALED BY DE NOVO GENOME PROJECT DATA

Á. Tóth[a], E. Baka[a], Sz. Luzics[a], I. Bata-Vidács[a], I. Nagy[b,c], B. Bálint[c], R. Herczeg[c], F. Olasz[d], T. Wilk[d], T. Nagy[d], B. Kriszt[e], I. Nagy[f] and J. Kukolya[a*]

[a]Department of Applied and Environmental Microbiology, Research Institute of Agro-Environmental Sciences, National Agricultural Research and Innovation Centre, H-1022 Budapest, Herman Ottó út 15. Hungary
[b]Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, H-6726 Szeged, Temesvári krt. 62. Hungary
[c]SeqOmics Biotechnology Ltd., H-6782 Mórahalom, Vállalkozók útja 7. Hungary
[d]Laboratory of Microbiology; Agricultural Biotechnology Institute, National Agricultural Research and Innovation Centre, H-2100 Gödöllő, Szent-Györgyi Albert utca 4. Hungary
[e]Department of Environmental Safety and Ecotoxicology, Institute of Aquaculture and Environmental Safety, Faculty of Agricultural and Environmental Sciences, Szent István University, H-2100 Gödöllő, Páter Károly utca 1. Hungary
[f]Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18., D-82152 Martinsried. Germany

(Received: 16 September 2015; accepted: 20 January 2016)

Thermobifidas are thermophilic, aerobic, lignocellulose decomposing actinomycetes. The *Thermobifida* genus includes four species: *T. fusca*, *T. alba*, *T. cellulosilytica*, and *T. halotolerans*. *T. fusca* YX is the far best characterized strain of this taxon and several cellulases and hemicellulases have been cloned from it for industrial purposes targeting paper industry, biofuel, and feed applications. Unfortunately, sequence data of such enzymes are almost exclusively restricted to this single species; however, we demonstrated earlier by zymography that other *T. alba* and *T. cellulosilytica* strains encode the same enzyme sets. Recently, the advances in whole genome sequencing by the use of next generation genomics platforms accelerated the selection process of valuable hydrolases from uncharacterized bacterial species for cloning purposes. For this purpose *T. cellulosilytica* TB100^T type strain was chosen for de novo genome sequencing. We have assembled the genome of *T. cellulosilytica* strain TB100^T into 168 contigs and 19 scaffolds, with reference length of 4 327 869 bps, 3 589 putative coding sequences, 53 tRNAs, and 4 rRNAs. The analysis of the annotated genome revealed the existence of 27 putative hydrolases belonging to 14 different glycoside hydrolase (GH) families. The investigation of identified, cloned, and heterologously multiple cellulases, mannanases, xylanases, and amylases may result in industrial applications beside gaining useful basic research related information.

Keywords: de novo genome project, glycoside hydrolyses, *Thermobifida,* cellulases, hemicellulases, amylases

Four species, *Thermobifida fusca*, *T. alba* (Zhang et al., 1998), *T. cellulosilytica* (Kukolya et al., 2002), and *T. halotolerans* (Yang et al., 2008), constitute the *Thermobifida* genus. Several cellulases and hemicellulases have been cloned from the best characterized *T. fusca* YX strain. This actinomycete utilizes various plant cell wall polymers, including cellulose as the major carbon source, and secretes multiple cellulases: four endoglucanases (Cel9B, Cel6A, Cel5A, Cel5B), two exoglucanases (Cel6B and Cel48A), and an endo/exoglucanase (Cel9A), which have been characterized in detail (Ghangas & Wilson, 1988; Irwin et al., 1998; Lao et

* To whom correspondence should be addressed.
Phone: +36 1 796-0400/186; fax: +36 1 796-0449; e-mail: j.kukolya@cfri.hu

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purposes, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated.

0139–3006 © 2017 *The Author(s)*

al., 1991; JUNG et al., 1993; POSTA et al., 2004). Unfortunately, sequence data of these enzymes are almost exclusively restricted to this single species. The complete genome-sequence of *T. fusca* strains were published in 2007 and 2013 (LYKIDIS et al., 2007; TÓTH et al., 2013). Earlier we demonstrated by zymography that *T. cellulosilytica* strains encode the same cellulase enzyme set as *T. fusca* TM51 (KUKOLYA et al., 2002). Despite the high scientific and industrial potential of hydrolases of *T. cellulosilytica*, there is only one paper dealing with a cutinase from this species (HERRERO et al., 2013).

For facilitating the research on *T. cellulosilytica* plant polysaccharide degrading enzyme system, the full genome sequencing seems the most appropriate and fastest approach. The Genomes OnLine Database (GOLD: www.genomesonline.org) (REDDY et al., 2015) currently holds 47 509 bacterial genome projects, 3948 of them being completed with closed genomes, while 22 739 of them are completed yielding permanent draft. It is worth noting how small fraction of the genome projects reaches complete closure. Despite the massive simplification and cost reduction offered by NGS technologies, assembling a bacterial genome into completely closed chromosome(s) is still a tedious work. Large repetitive elements (rRNA clusters, prophage regions, transposable elements, etc.) are difficult if not impossible regions for the assembler software to accurately resolve that results gaps in the draft genome. Finishing a draft to completely closed sequences requires classic molecular (low throughput) biology techniques (PCR) and multiple rounds of capillary electrophoresis sequencing.

Our aim was to generate a good quality, scaffolded draft genome of *Thermobifida cellulosilytica* strain that can be later subjected to automated annotation yielding a comprehensive list of biochemical features of the organism. As the main feature of thermobifidas is the outstanding cellulose degradation capacity, we focused on the identification of the main plant polysaccharide degrading enzyme complexes, namely cellulases, hemicellulases, and amylases.

1. Materials and methods

1.1. Chemicals

Unless otherwise indicated, all chemicals herein used were of analytical-grade and purchased from Sigma-Aldrich Ltd. (Budapest, Hungary).

1.2. Microorganisms

Ten *Thermobifida* strains were used in this study. *Thermobifida fusca* TM51, TB107, *T. alba* K51, K52, and *T. cellulosilytica* K21, TB100^T, TB108, and TB110 were isolated earlier from the hot region of manure compost (KUKOLYA et al., 1997). *T. fusca* 27730^T and *Thermobifida alba* 43795^T were purchased from American Type Culture Collection (ATTC) and Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), respectively.

1.3. Endoglucanase zymography for selecting T. cellulosilytica *strain for de novo genome project*

1.3.1. Production of crude cellulase enzyme samples for zymography. Thermobifida strains were grown on basal medium ($NaNO_3$, 1.0 g; KCl, 0.3 g; $MgSO_4$ •H_2O, 0.5 g; K_2HPO_4, 1.0 g; yeast extract, 0.5 g; peptone, 0.5 g; distilled water, 1000 ml, pH 7.6) containing 0.2% carboxymethyl cellulose. For endoglucanase enzyme production, liquid cultures were shaken

at 220 r.p.m. for 72 h at 50 °C. After centrifugation, supernatants were treated with EtOH, reaching 40% ethanol concentration. The precipitates – representing the crude enzyme fraction – were separated by centrifugation at 15 000×*g* for 15 min at 4 °C.

1.3.2. SDS-PAGE zymography. SDS-polyacrylamide slab gels (0.75 mm thick) were prepared according to the method of O'Farrell (1975). The 10% acrylamide separating gel contained 0.1% CMC as substrate to test endoglucanase activity (Posta et al., 2004). Aliquots of 5 µl ethanol precipitated crude enzyme samples were boiled in loading buffer for 5 min. Gels were renatured after separation by washing twice with isopropanol solution containing (25%, V/V) phosphate buffer (0.1 M, pH 7.0) at room temperature (RT). After this, two subsequent washing steps with phosphate buffer (0.1 M, pH 7.0) were applied and gels were incubated at 50 °C for 30 min and were stained for visualizing active enzyme bands with 1% Congo-red solution.

1.4. De novo genome sequencing

Genomic DNA from *T. cellulosilytica* TB100^T was prepared as previously described (Béki et al., 2003). Mate-pair genomic library kit (Illumina) was used to construct a genomic library from *T. cellulosilytica* genomic DNA with template sizes between 3 and 6 kb. Samples were sequenced on an Illumina MiSeq platfom using 2×250 cycles and V2 sequencing chemistry. Mate-pair junction adapter was identified and removed from the raw reads according to the manufacturer's recommendations (Illumina). De novo assembly of the processed MP reads was carried out with Mira V 4.0.2 (Chevreux et al., 1999) and contigs were scaffolded using SSpace 3.0. (Boetzer et al., 2011); draft genome was annotated using RAST (Aziz et al., 2008).

1.5. Identification and characterization of polysaccharide hydrolases

For characterizing the main plant polysaccharide degrading enzyme sets (cellulases, hemicellulases, and amylases) from the annotated genome, sequences were analysed by using the BLAST server (Madden et al., 1996). Amino acid sequence and domain structure of the selected glycoside hydrolases were determined by Swiss-Prot, EMBL and NCBI database queries and by using the Pfam (Bateman et al., 2002) and InterPro (Mitchell et al., 2015) bioinformatics servers.

2. Results and discussion

2.1. Endoglucanase zymography to select T. cellulosilytica *strain for de novo genome project*

Crude cellulases of ten different *Thermobifida* strains belonging to *T. fusca, T. alba,* and *T. cellulosilytica* were investigated by endoglucanase zymography. As clearly can be seen in Figure 1, endoglucanases produced by *Thermobifida* strains are extremely heat stable, as these enzymes remained active even after 5 min boiling in loading buffer. According to the band pattern, the three *Thermobifida* species produce unique and slightly different endoglucanases. The number of endoglucanase bands is higher (5–7) in the case of *T. alba* and *T. fusca* than for *T. cellulosilytica* (3–6). According to the multiplicity of endoglucanase zymogram, *T. cellulosilytica* TB100^T was chosen for the de novo genome project.

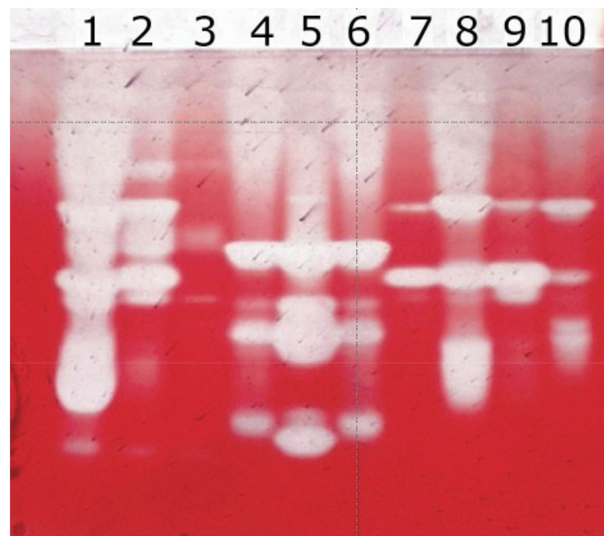

Fig. 1. SDS-PAGE endoglucanase zymogram of *Thermobifida* strains. Active endoglucanases were visualized by Congo-red staining. Samples: 1: *T. alba* 43795^T; 2: K51; 3: K52; 4: *T. fusca* 27730^T; 5: TM51; 6: TB107; 7: *T. cellulosilytica* K21; 8: *T. cellulosilytica* TB100^T; 9: TB108; 10: TB110

2.2. De novo genome sequencing

We have assembled the genome of *T. cellulosilytica* strain TB100^T into 168 contigs and 19 scaffolds, with reference length of 4 327 869 bps, 3589 putative coding sequences, 53 tRNAs, and 4 rRNAs. The analysis of the annotated genome revealed the existence of 27 putative hydrolases belonging to 14 different glycoside hydrolase (GH) families.

This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession numbers: BioProject; PRJNA289053, BioSample; SAMN03840822, NCBI Accession; LGEM00000000.

2.2.1. Cellulases. The cellulase enzyme system of *T cellulosilytica* (Fig. 2) found in our genome project is identical to the well-characterized glycoside hydrolases of *T. fusca* (Wilson, 2004). Characteristic features of the enzymes are summarized in Table 1. Four glycoside hydrolase enzyme sets were found that belonged to four independent families. Each of the enzyme sets consisted of 4 endoglucanases (endo-β-1,4-glucanase/cellulase, EC 3.2.1.4), Cel5A, Cel5B, Cel6A, Cel9B (Irwin et al., 1993; Jung et al., 1993; Zhang & Wilson, 1997; Posta et al., 2004), one processive endoglucanase, Cel 9A (Sakon et al., 1997), and two cellulose chain reducing cellobiohydrolases. One of them (Cel 48A, EC 3.2.1.176) is acting on the reducing, the other one, Cel6B (cellobiohydrolase, EC 3.2.1.91), is on the non-reducing end (Zhang et al., 1995; Barr et al., 1996).

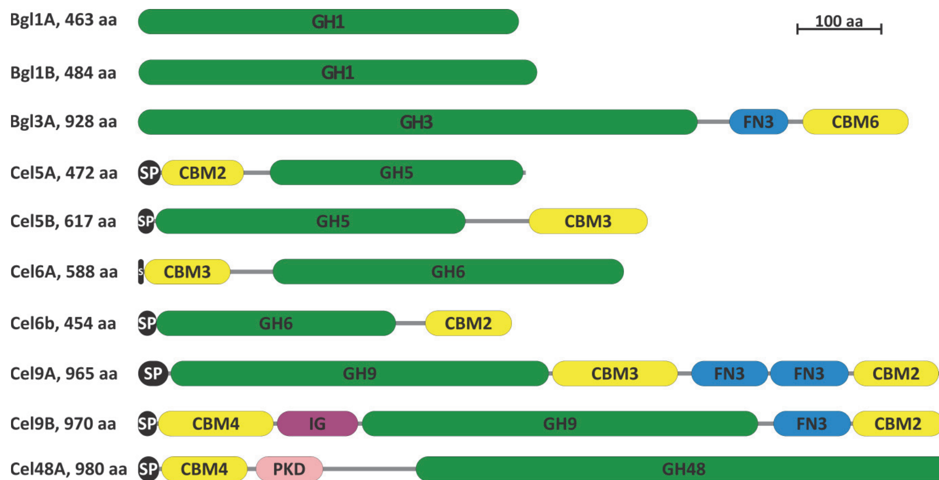

Fig. 2. Domain structure of cellulases encoded by *T. cellulosilytica* TB100[T]. SP: signal peptide; CBM: carbohydrate modules; FN3: fibronectin type domain; IG: immune globulin like domain; PKD: polycystic kidney disease protein like domain

A common feature of extracellular enzymes of *T. cellulosilytica* is the complex domain structure, where both catalytic domain and hydrophobic substrate binding modules, like CBD 2, 3, and 6 types, FN3 and IG modules, and PKD domains are present (Figs 2–5). Additionally, linker sequences with characteristic motives can be found between domains of cellulase enzymes except enzymes of the GH9 family (Table 1). These might have a role in post-translational modifications providing cleavage sites for specific proteases. By this mechanism there is a possibility for the representation of "only catalytic site" enzymes that can hydrolyse shorter, water soluble polymers/oligosaccharides. Such maturation mechanism can be responsible for the increased number of active bands on the zymogram (Fig. 1), since more bands are visualized than could be expected after the number of endoglucanases encoded on the genome.

There were also two intracellular GH1 family beta-glucosidases (EC 3.2.1.21) and a GH3 family hypothetical glucan 1,3-β-glucosidase (EC 3.2.1.58) found. Interestingly, the latter enzyme seems to be intracellularly located in spite of encoding for complex polysaccharide binding domains (FN3 and CBD6).

2.2.2. Hemicellulases. The identified hemicellulases from the annotated genome are shown in Figure 3, their features are summarized in Table 1. The hemicellulases of *T. cellulosilytica* show total homology to the xylanases, mannanases, and xyloglucanases identified in *T. fusca* genome project. The polysaccharide binding domains of this enzyme group are less diverse compared to the cellulases: the extracellular enzymes belong to CBM2 type without exception. Interestingly, a TAT signal sequence of GH74 xyloglucanase enzyme (EC 3.2.1.151), which refers to the presence of transport system specific for *Streptomyces*, was identified. This enzyme was formerly described in *T. fusca* (Irwin et al., 2003). The mannanases among the hemicellulases consist of an intracellular GH2 mannosidase and an extracellular GH5 endomannanase. From *T. fusca* only the mannosidase enzyme was

Table 1. Features of glycosyde hydrolases encoded by *Thermobifida cellulosilytica* TB100^{T}

	GH family	CBD	Signal sequence	Linker motif	Protein length (aa)	Molecular mass (Da)	Isoelectric point	Accession number	Publication
Cellulase									
Beta-glucosidase, Bgl1A	GH1	no	no	no	463	51682	5.25	AC529_00400	no
Beta-glucosidase, Bgl1B	GH1	no	no	no	484	53255	4.85	AC529_14095	Spiridonov & Wilson (2001)
Beta-glucosidase, Bgl3A	GH3	CBM6, FN3	no	no	928	99893	5.06	AC529_16850	no
Endoglucanase, Cel5A	GH5	CBM2	1–36	(TDPG)4	472	50124	5.23	AC529_18225	Irwin et al. (1993)
Endoglucanase, Cel5B	GH5	CBM3	1–28	(PTD)4(PAD) (PTD)5(PAD)	618	68352	4.51	AC529_03550	Posta et al. (2004)
Endoglucanase, Cel6A	GH6	CBM2	1–16	(DGGN)10	588	62012	4.30	AC529_08125	Zhang & Wilson (1997)
Exoglucanase, Cel6B	GH6	CBM2	1–31	(GNPNNP)2 (GNPDNP) (GNPNNP)	454	46621	5.77	AC529_09675	Spezio et al. (1993), Irwin et al. (1993), Taylor et al. (1995)
Endo/exo-glucanase, Cel9A	GH9	CBM3, FN3, CBM2	1–45 (TAT)	no	965	103866	4.67	AC529_10315	Sakon et al. (1997)
Endoglucanase, Cel9B	GH9	CBM4_9, IG–like, FN3, CBM2	1–18	no	957	102496	4.14	AC529_00410	Jung et al. (1993)
Exoglucanase, Cel48A	GH48	CBM2	1–34	(SGSS)DLTV (SGSS)	981	106345	4.35	AC529_09005	Irwin et al. (2000)
Hemicellulase									
Beta-mannosidase, Man2	GH2	SBD, IG–like	no	no	831	92633	4.91	AC529_18295	Béki et al. (2003)
Endomannanase, Man5	GH5	CBM2	1–28	(DPGT)6	452	47541	4.76	AC529_18220	Kumagai et al. (2011), Kumagai et al. (2012)
Xyloglucanase, Xeg74	GH74	CBM2	1–49 (TAT)	(EEP)9	919	97875	4.29	AC529_00500	Irwin et al. (2003)

Table 1. continued

	GH family	CBD	Signal sequence	Linker motif	Protein length (aa)	Molecular mass (Da)	Isoelectric point	Accession number	Publication
Endoxylanase, Xyl10A	GH10	CBM2	1–35	no	474	50546	4.49	AC529_07560	
Endoxylanase, Xyl10B	GH10	no	1–27	(GPGH)5 (GPGQ)2 (GPGH)3 GPG	424	46330	4.96		Kim et al. (2004), Moraïs et al. (2011)
Endoxylanase, Xyl11	GH11	CBM2	1–23	(NNGGG)4	327	34815	8.98		Ghangas et al. (1989), Irwin et al. (1994), Moraïs et al. (2011)
Beta-xylosidase, Xyl43	GH43	CAL	no	no	546	61846	5.88	AC529_00480	Morais et al. (2012) Fekete & Kiss (2013)
Amylase, chitinase									
Alpha-amylase, Amy13A	GH13	CBM20	1–33	no	607	64772	4.71	AC529_14300	Yang & Liu (2007)
Alpha-amylase, Amy13B	GH13	no	no	no	654	73439	5.79		
Alpha-amylase, Amy13C	GH13	no	no	no	651	72469	5.76	AC529_00690	
Alpha-amylase, Amy13D	GH13	no	no	no	544	60737	5.18	AC529_09800	
Alpha-amylase, Amy13E	GH13	no	no	no	544	60086	4.79	AC529_18630	
Alpha-amylase, Amy13F	GH13	no	no	no	775	84655	5.94	AC529_00365	
Alpha-amylase, Amy13G	GH13	no	no	no	359	42104	4.75	AC529_00695	Wei et al. (2004)
Alpha-amylase, Amy13H	GH13	no	no	no	648	72789	5.74	AC529_06800	
Amylomaltase, Amy77	GH77	no	no	no	708	77899	6.21	AC529_10835	
Chitinase, Chi	GH18	no	no	no	416	46151	4.54	AC529_18825	

described so far (BÉKI et al., 2003), the endomannanase has not been published yet. *T. cellulosilytica* encodes the following xylanases: two GH10 and one GH11 extracellular endoxylanases (endo-1,4-β-xylanase (EC 3.2.1.8) and one GH43 intracellular beta-xylosidase (EC 3.2.1.37)). Although the first *Thermobifida* crystallized hydrolase, a GH11 endoxylanase, was published in 1996 (HILGE et al., 1996) and after that a GH10 endoxylanase from *T. alba* was published (BLANCO et al., 1997), the Xyl10B enzyme has not yet been described.

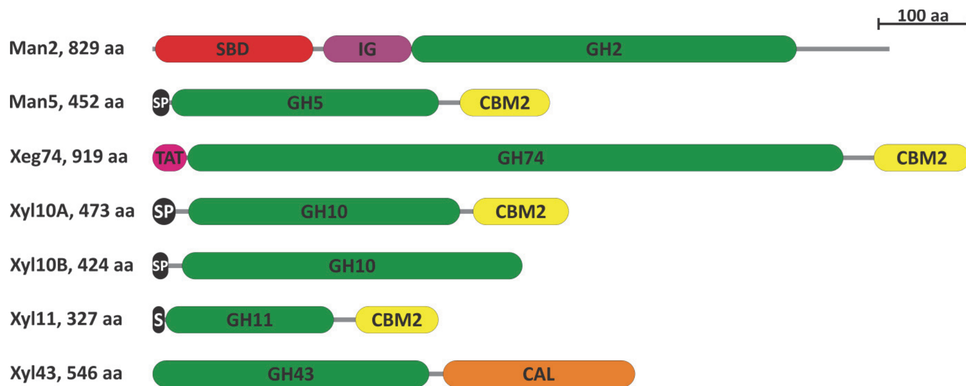

Fig. 3. Domain structure of hemicellulases encoded by *T. cellulosilytica* TB100^T. SP: signal peptide; TAT: streptomyces transport signal; CBM: carbohydrate modules; IG: immune globulin like domain; CAL: carbohydrate binding; SBD: sugar binding domain

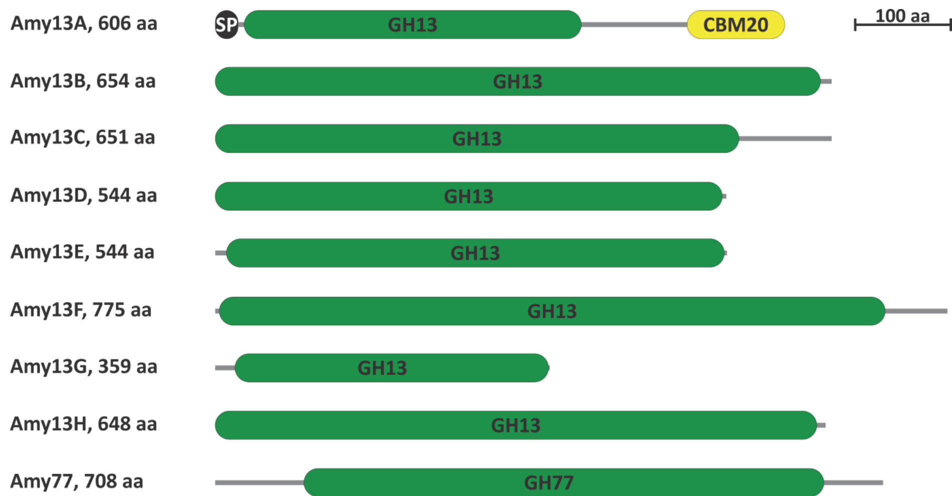

Fig. 4. Domain structure of amylases encoded by *T. cellulosilytica* TB100^T. SP: signal peptide; CBM: carbohydrate module

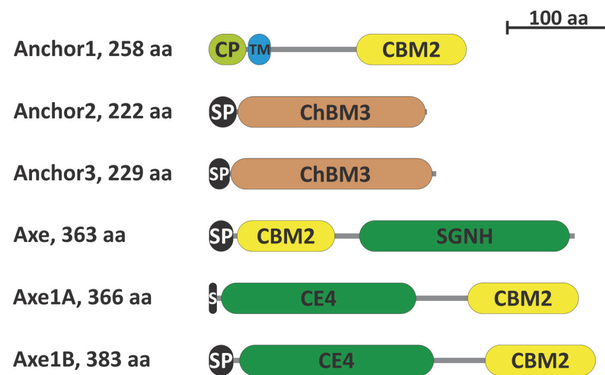

Fig. 5. Domain structure of proteins harbouring carbohydrate binding modules produced by *T. cellulosilytica* TB100^{T}. SP: signal peptide; CP: cytoplasmic module; TM: transmembrane protein; ChBM: chitin binding module; CBM: carbohydrate module; SGNH: hydrolase domain; CE: carbohydrate esterase domain

2.2.3. Amylases. The amylase system of *T. cellulosilytica* consists of 9 identified enzymes, one of which (amylomaltase or 4-α-glucanotransferase, EC 2.4.1.25) is from the family GH77, and 8 are from the family GH13. The single extracellular member of the system is enzyme Amy13A, bearing a CBM20 amylose binding domain N-terminally (Fig. 4). The high number of intracellular GH13 enzymes presumes a distinct function, which is supported by the extremely wide substrate spectrum, the mechanism of actions, and prevalence of this hydrolase family among bacteria (according to the Cazy database, 24605 genes coding this enzyme has been identified to this day): α-amylase (EC 3.2.1.1); pullulanase (EC 3.2.1.41); cyclomaltodextrin glucanotransferase (EC 2.4.1.19); cyclomaltodextrinase (EC 3.2.1.54); trehalose-6-phosphate hydrolase (EC 3.2.1.93); oligo-α-glucosidase (EC 3.2.1.10); maltogenic amylase (EC 3.2.1.133); neopullulanase (EC 3.2.1.135); α-glucosidase (EC 3.2.1.20); maltotetraose-forming α-amylase (EC 3.2.1.60); isoamylase (EC 3.2.1.68); glucodextranase (EC 3.2.1.70); maltohexaose-forming α-amylase (EC 3.2.1.98); maltotriose-forming α-amylase (EC 3.2.1.116); branching enzyme (EC 2.4.1.18); trehalose synthase (EC 5.4.99.16); 4-α-glucanotransferase (EC 2.4.1.25); maltopentaose-forming α-amylase (EC 3.2.1.-); amylosucrase (EC 2.4.1.4); sucrose phosphorylase (EC 2.4.1.7); malto-oligosyltrehalose trehalohydrolase (EC 3.2.1.141); isomaltulose synthase (EC 5.4.99.11); malto-oligosyltrehalose synthase (EC 5.4.99.15); amylo-α-1,6-glucosidase (EC 3.2.1.33); α-1,4-glucan: phosphate α-maltosyltransferase (EC 2.4.99.16).

The research on *Thermobifida* amylases is a neglected area, only two enzymes are mentioned in the literature: GH77 amylomaltase of *Thermobifida fusca* (Yang & Liu, 2007) and a GH13 trehalose synthase (Wei et al., 2004).

2.2.4. CBM containing proteins. Five extracellular proteins containing carbohydrate binding modules could be identified in the annotated genome. Three of them are acetyl xylan esterases (EC 3.1.1.72) bearing one module of CBM2 in C- or N-terminal positions. There are also three proteins without catalytic domain, the first bearing the CBM2 module, bound to cell surface, indicated by its transmembrane region. The function of this protein could be similar to the cellulose anchoring protein of *Streptomyces reticuli* (Walter et al., 1998). The other two show chitin binding module homology, the role of these is yet totally unknown in thermobifidas (Fig. 5).

3. Conclusions

Thermobifida cellulosilytica strain TB100^T was chosen for de novo genome project based on the highest number of isoenzyme bands on endoglucanase zymogram. Strain TB100^T has a complex endoglucanase isoenzyme pattern like *T. fusca* and *T. alba*. According to the band pattern, the three *Thermobifida* species produce unique and slightly different endoglucanases. The good resolution of the endoglucanase zymogram enables the identification of unknown thermobifida isolates. The number of endoglucanase bands is higher (5–7) than the formerly described endoglucanases (Cel5A, Cel5B, Cel6A, Cel9A) of *T. fusca* YX strain (Wilson, 2004). The probable explanation could be found in the post-translational modifications of the modular enzymes, during which a protease splits the cellulases into a catalytic and a cellulose-binding domain, increasing the number of isoenzymes.

Based on the chosen *T. cellulosilytica* TB100^T de novo genome project, the genome shows great similarity to the two already available genomes of *T. fusca*. The assembled genome of *T. cellulosilytica* strain TB100^T contains 168 contigs and 19 scaffolds, with reference length of 4 327 869 bps. The *T. cellulosilytica* TB100^T genome size is larger by 0.7 Mbp than the *T. fusca* TM51 (Tóth et al., 2013) or YX (Lykidis et al., 2007) genomes. Despite the larger genome size, the complexity of the glycoside hydrolase system, taking part in the degradation of plant polysaccharides of TB100^T strain, is identical to *T. fusca* YX (Wilson, 2004) and TM51 strains (Tóth et al., 2013), additional enzymes were not found. The analysis of the annotated genome revealed the existence of 27 putative hydrolases belonging to 14 different glycoside hydrolase (GH) families (Cantarel et al., 2009). GH13 is the largest glycoside hydrolase family in TB100^T, with eight enzymes predicted to exhibit mainly dextran- and starch-degrading functions. Thermobifidas may become important industrial strains due to their thermostable and robust polysaccharide-degrading enzymes (Dhawan & Kaur, 2007; Yang et al., 2007). The 83–87 percent amino acid sequence similarity of the glycoside hydrolases of *Thermobifida* species seems to be low enough to exhibit differences in the biochemical properties compared to other *T. fusca* enzymes. This can be clearly seen in the endomannanase zymograms of thermobifidas, where endomannanases of *T. cellulosilytica* strains proved to be the least heat stable. The genome project sequence of *T. cellulosilytica* makes the cloning of homologous, known enzymes of fusca-origin or unknown glycoside hydrolase enzymes not yet described from thermobifidas possible. Thus, enzymes encoded by *Thermobifida cellulosilytica* TB100^T may play a key role not only in pure science but in further industrial applications, like lignocellulose-based ethanol-producing projects, prebiotics production, and as sources for feed additives.

*

This work was supported by the Research Centre of Excellence-9878/2015/FEKUT project. István Nagy was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko, O. (2008): The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* doi: 10.1186/1471-2164-9-75

Barr, B.K., Hsieh, Y.L., Ganem, B. & Wilson, D.B. (1996): Identification of two functionally different classes of exocellulases. *Biochemistry*, *35*, 586–592.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. & Sonnhammer, E.L. (2002): The Pfam protein families database. *Nucleic Acids Res*., *30*, 276–280.

Béki, E., Nagy, I., Vanderleyden, J., Jager, S., Kiss, L., Fulop, L., Hornok, L. & Kukolya, J. (2003): Cloning and heterologous expression of a beta-D-mannosidase (EC 3.2.1.25) encoding gene from *Thermobifida fusca* TM51. *Appl. Environ. Microb*., *69*, 1944–1952.

Blanco, J., Coque, J.J., Velasco, J. & Martin, J.F. (1997): Cloning, expression in *Streptomyces lividans* and biochemical characterization of a thermostable endo-beta-1,4-xylanase of *Thermomonospora alba* ULJB1 with cellulose-binding ability. *Appl. Microbiol. Biot., 48*, 208–217.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D. & Pirovano, W. (2011): Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, *27*, 578–579.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009): The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res*., *37*, D233–D238.

Chevreux, B., Wetter, T. & Suhai, S. (1999): Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics*, *99*, 45–56.

Dhawan, S. & Kaur, J. (2007): Microbial mannanases: an overview of production and applications. *Crit. Rev. Biotechnol*., *27*, 197–216.

Fekete, C.A. & Kiss, L. (2013): A new approach in the active site investigation of an inverting β-D-xylosidase from *Thermobifida fusca* TM51. *Protein J.*, *32*, 97–105. doi: 10.1007/s10930-013-9463-8.

Ghangas, G.S. & Wilson, D.B. (1988): Cloning of the *Thermomonospora fusca* endoglucanase E2 gene in *Streptomyces lividans*: Affinity purification and functional domains of the cloned gene product. *Appl. Environ. Microbiol*., *54*, 2521–2526.

Ghangas, G.S., Hu, Y.J. & Wilson, D.B. (1989): Cloning of a *Thermomonospora fusca* xylanase gene and its expression in *Escherichia coli* and *Streptomyces lividans*. *J. Bacteriol*., *171*, 2963–2969.

Herrero, E., Ribitsch, D., Dellacher, A., Zitzenbacher, S., Marold, A., Steinkellner, G., Gruber, K., Schwab, H. & Guebitz, G.M. (2013): Surface engineering of a cutinase from *Thermobifida cellulosilytica* for improved polyester hydrolysis. *Biotechnol. Bioeng., 110*, 2581–2590.

Hilge, M., Gloor, S., Winterhalter, K., Zimmermann, W. & Piontek, K. (1996): Crystallization and preliminary crystallographic analysis of two beta-mannanase isoforms from *Thermomonospora fusca* KW3. *Acta Crystallogr. D., 52*, 1224–1225.

Irwin, D.C., Jung, E.D. & Wilson, D.B. (1994): Characterization and sequence of a *Thermomonospora fusca* xylanase. *Appl. Environ. Microb*., *60*, 763–770.

Irwin, D.C., Shin, D.H., Zhang, S., Barr, B.K., Sakon, J., Karplus, P.A. & Wilson, D.B. (1998): Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J. Bacteriol*., *180*, 1709–1714.

Irwin, D.C., Cheng, M., Xiang, B., Rose, J.K. & Wilson, D.B. (2003): Cloning, expression and characterization of a family-74 xyloglucanase from *Thermobifida fusca*. *Eur. J. Biochem*., *270*, 3083–3091.

Irwin, D.C., Spezio, M., Walker, L.P. & Wilson, D.B. (1993): Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. *Biotechnol. Bioeng.*, *42*, 1002–1013.

Irwin, D.C., Zhang, S. & Wilson, D.B. (2000): Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from *Thermobifida fusca*. *Eur. J. Biochem*., *267*, 4988–4997.

Jung, E.D., Lao, G., Irwin, D., Barr, B.K., Benjamin, A. & Wilson, D.B. (1993): DNA sequences and expression in *Streptomyces lividans* of an exoglucanase gene and an endoglucanase gene from *Thermomonospora fusca*. *Appl. Environ. Microb*., *59*, 3032–3043.

Kim, J.H., Irwin, D. & Wilson, D.B. (2004): Purification and characterization of *Thermobifida fusca* xylanase 10B. *Can. J. Microbiol., 50*, 835–843.

Kukolya, J., Dobolyi, C. & Hornok, L. (1997): Isolation and identification of thermophilic cellulolytic actinomycetes. *Acta Phytopathol. Hun*., *32*, 97–107.

Kukolya, J., Nagy, I., Láday, M., Tóth, E., Oravecz, O., Márialigeti, K. & Hornok, L. (2002): *Thermobifida cellulolytica* sp. nov., a novel lignocellulose-decomposing actinomycete. *Int. J. Syst. Evol. Micr., 52*, 1193–1199.

KUMAGAI, Y., KAWAKAMI, K., MUKAIHARA, T., KIMURA, M. & HATANAKA, T. (2012): The structural analysis and the role of calcium binding site for thermal stability in mannanase. *Biochimie*, *94*, 2783–2790. doi: 10.1016/j.biochi.2012.09.012. Epub 2012 Sep 23.

KUMAGAI, Y., USUKI, H., YAMAMOTO, Y., YAMASATO, A., ARIMA, J., MUKAIHARA, T. & HATANAKA, T. (2011): Characterization of calcium ion sensitive region for β-mannanase from *Streptomyces thermolilacinus*. *Biochim. Biophys. Acta*, *1814*, 1127–1133. doi: 10.1016/j.bbapap.2011.04.017. Epub 2011 May 12.

LAO, G., GHANGAS, G.S., JUNG, E.D. & WILSON, D.B. (1991): DNA sequences of three beta-1,4-endoglucanase genes from *Thermomonospora fusca*. *J. Bacteriol., 173,* 3397–3407.

LYKIDIS, A., MAVROMATIS, K., IVANOVA, N., ANDERSON, I., LAND, M., DIBARTOLO, G., MARTINEZ, M., LAPIDUS, A., LUCAS, S., COPELAND, A., RICHARDSON, P., WILSON, D.B. & KYRPIDES, N. (2007): Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. *J. Bacteriol.*, *189*, 2477–2486.

MADDEN, T.L., TATUSOV, R.L. & ZHANG, J. (1996): Applications of network BLAST server. *Methods Enzymol*., *266*, 131–141.

MITCHELL, A., CHANG, H.Y., DAUGHERTY, L., FRASER, M., HUNTER, S., LOPEZ, R., MCANULLA, C., MCMENAMIN, C., NUKA, G., PESSEAT, S., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., RATO, C., YONG, S.Y., BATEMAN, A., PUNTA, M., ATTWOOD, T.K., SIGRIST, C.J., REDASCHI, N., RIVOIRE, C., XENARIOS, I., KAHN, D., GUYOT, D., BORK, P., LETUNIC, I., GOUGH, J., OATES, M., HAFT, D., HUANG, H., NATALE, D.A., WU, C.H., ORENGO, C., SILLITOE, I., MI, H., THOMAS, P.D. & FINN, R.D. (2015): The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res*., *43*, D213–D221.

MORAÏS, S., BARAK, Y., HADAR, Y., WILSON, D.B., SHOHAM, Y., LAMED, R. & BAYER, E.A. (2011): Assembly of xylanases into designer cellulosomes promotes efficient hydrolysis of the xylan component of a natural recalcitrant cellulosic substrate. *MBio*., *2*, 233–211. doi: 10.1128/mBio.00233-11. Print 2011.

MORAÏS, S., SALAMA-ALBER, O., BARAK, Y., HADAR, Y., WILSON, D.B., LAMED, R., SHOHAM, Y. & BAYER, E.A. (2012): Functional association of catalytic and ancillary modules dictates enzymatic activity in glycoside hydrolase family 43 β-xylosidase*. J. Biol. Chem*., *287*, 9213–9221. doi: 10.1074/jbc.M111.314286. Epub 2012 Jan 23.

O'FARRELL, P.H. (1975): High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, *250*, 4007–4021.

POSTA, K., BÉKI, E., WILSON, D.B., KUKOLYA, J. & HORNOK, L. (2004): Cloning, characterization and phylogenetic relationships of cel5B, a new endoglucanase encoding gene from *Thermobifida fusca*. *J. Basic Microb*., *44*, 383–399.

REDDY, T.B.K., THOMAS, A.D., STAMATIS, D., BERTSCH, J., ISBANDI, M., JANSSON, J., MALLAJOSYULA, J., PAGANI, I., LOBOS, E.A. & KYRPIDES, N.C. (2015): The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleid Acids Res*., *43*, 1099–1106.

SAKON, J., IRWIN, D., WILSON, D.B. & KARPLUS, P.A. (1997): Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat. Struct. Biol.*, *4*, 810–818.

SPEZIO, M., WILSON, D.B. & KARPLUS, P.A. (1993): Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry*, *32*, 9906–9916.

SPIRIDONOV, N.A. & WILSON, D.B. (2001): Cloning and biochemical characterization of BglC, a beta-glucosidase from the cellulolytic actinomycete *Thermobifida fusca*. *Curr Microbiol*., *42*, 295–301.

TAYLOR, J.S., TEO, B., WILSON, D.B. & BRADY, J.W. (1995): Conformational modeling of substrate binding to endocellulase E2 from *Thermomonospora fusca*. *Protein Eng*., *8*, 1145–1152.

TÓTH, Á., BARNA, T., NAGY, I., HORVÁTH, B., TÁNCSICS, A., KRISZT, B., BAKA, E., FEKETE, CS. & KUKOLYA, J. (2013): Draft genome sequence of the lignocellulose decomposer *Thermobifida fusca* strain TM51. *Genome announcements*, *1*(4), e00482-13.

WALTER, S., WELLMANN, E. & SCHREMPF, H. (1998): The cell wall-anchored *Streptomyces reticuli* avicel-binding protein (AbpS) and its gene. *J. Bacteriol.*, *180*, 1647–1654.

WEI, Y.T., ZHU, Q.X., LUO, Z.F., LU, F.S., CHEN, F.Z., WANG, Q.Y., HUANG, K., MENG, J.Z., WANG, R. & HUANG, R.B. (2004): Cloning, expression and identification of a new trehalose synthase gene from *Thermobifida fusca* genome. *Acta Biochim. Biophys. Sin.* (Shanghai), *36*, 477–484.

WILSON, D.B. (2004): Studies of *Thermobifida fusca* plant cell wall degrading enzymes. *Chem. Rec*., *4*, 72–82.

YANG, C.H., YANG, S.F. & LIU, W.H. (2007): Production of xylooligosaccharides from xylans by extracellular xylanases from *Thermobifida fusca*. *J. Agr. Food Chem*., *55*, 3955–3959.

YANG, C.H. & LIU, W.H. (2007): Cloning and characterization of a maltotriose-producing alpha-amylase gene from *Thermobifida fusca*. *J. Ind. Microbiol. Biot.*, *34*, 325–330.

Yang, L.L., Tang, S.K., Zhang, Y.Q., Zhi, X.Y., Wang, D., Xu, L.H. & Li, W.J. (2008): *Thermobifida halotolerans* sp. nov., isolated from a salt mine sample, and emended description of the genus *Thermobifida*. *Int. J. Syst. Evol. Microb*., *58*, 1821–1825.

Zhang, S. & Wilson, D.B. (1997): Surface residue mutations which change the substrate specificity of *Thermomonospora fusca* endoglucanase E2. *J. Biotechnol*., *57*, 101–113.

Zhang, S., Lao, G. & Wilson, D.B. (1995): Characterization of a *Thermomonospora fusca* exocellulase. *Biochemistry*, *34*, 3386–3395.

Zhang, Z., Wang, Y. & Ruan, J. (1998): Reclassification of *Thermomonospora* and *Microtetraspora*. *Int. J. Syst. Bacteriol*., *48*, 411–422.