Novel Hyperthermophilic Crenarchaeon *Thermofilum adornatum* sp. nov. Uses GH1, GH3, and Two Novel Glycosidases for Cellulose Hydrolysis

Kseniya S. Zayulina[1], Tatiana V. Kochetkova[1], Ulyana E. Piunova[1,2], Rustam H. Ziganshin[3], Olga A. Podosokorskaya[1] and Ilya V. Kublanov[1]*

[1] Research Center of Biotechnology, Winogradsky Institute of Microbiology, Russian Academy of Sciences, Moscow, Russia,
[2] Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia,
[3] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

A novel hyperthermophilic, anaerobic filamentous archaeon, *Thermofilum adornatum* strain 1910b^T, is capable of growing with cellulose as its sole carbon and energy source. This strain was isolated from a terrestrial hot spring in Kamchatka, Russia. The isolate 1910b^T grew optimally at a temperature of 80°C and a pH of 5.5–6.0, producing cell-bound inducible cellulases. During genome analysis, genes, encoding various glycosidases (GHs) involved in oligo- and polysaccharide hydrolysis and genes for the fermentation of sugars were identified. No homologs of currently known cellulase families were found among the GHs encoded by the 1910b^T genome, suggesting that novel proteins are involved. To figure this out, a proteomic analysis of cells grown on cellulose or pyruvate (as a control) was performed. Both in-depth genomic and proteomic analyses revealed four proteins (Cel25, Cel30, Cel40, and Cel45) that were the most likely to be involved in the cellulose hydrolysis in this archaeon. Two of these proteins (Cel30 and Cel45) were hypothetical according to genome analysis, while the other two (Cel25 and Cel40) have GH3 and GH1 domains, respectively. The respective genes were heterologously expressed in *Escherichia coli* BL21 (DE3), and enzymatic activities of recombinant proteins were measured with carboxymethyl cellulose (CMC), Avicel and cellobiose as substrates. It was revealed that the Cel30 and Cel25 proteins were likely exoglucanases with side beta-glucosidase and endoglucanase activities, that Cel40 was a multifunctional glucanase capable of hydrolyzing beta-1,4-glucosides of various lengths, and that Cel45 was an endoglucanase with side exoglucanase activity. Taking into account that the cellulolytic activity of *T. adornatum* 1910b^T surface protein fractions was inducible, that recombinant Cel25 and Cel30 were much less active than Cel40 and Cel45, and that their gene expressions were (almost) non-induced by CMC, we suggest that Cel40 and Cel45 play a major role in the degradation of cellulose, while Cel25 and Cel30 act only as accessory enzymes.

Keywords: cellulase, endoglucanase, glycosidase, cellulose degradation, hyperthermophilic, *Crenarchaeota*, *Thermofilum*

INTRODUCTION

Cellulose is the most abundant organic polymer on Earth; this is why it is a significant part of Earth's carbon cycle (Maleki et al., 2016). Cellulases (endo-beta-1,4-glucanases) play a crucial role in the degradation of cellulose, and this makes them high-demand enzymes in biomass utilization industries. Endoglucanases are involved in many industrial processes, such as decreasing pulp viscosity in the paper and pulp industry and biopolishing cotton in the textile industry. They are also involved in 2nd generation biofuel production for the enzymatic pretreatment of cellulosic feedstock (Robb et al., 2007; Sims et al., 2010; Kang et al., 2014; Cabrera and Blamey, 2018; Robak and Balcerek, 2018) as well as other economic sectors. Thermostable cellulases offer several benefits in comparison with their mesophilic counterparts. High reaction temperatures decrease the risk of contamination, reduce viscosity, and increase the solubility of substrates, which increases the final yield of the end products (Haki and Rakshit, 2003; Egorova and Antranikian, 2005). Additionally, thermostable enzymes are highly resistant to denaturation and are highly stable during long-term storage, which lengthens both their operational life and their shelf life (Viikari et al., 2007; Peacock et al., 2013).

The following enzymes are the main actors in the hydrolysis of cellulose: endoglucanases/cellulases (3.2.1.4), enzymes that cut internal beta-1,4 glucosidic bonds and resulted in the formation of cellooligosaccharides or cellodextrins, which should be hydrolyzed further by exoglucanases/cellodextrinases (EC 3.2.1.74); beta-glucosidases (EC 3.2.1.21); cellobiose phosphorylases (EC 2.4.1.20); cellodextrin phosphorylases (EC 2.4.1.49); and a few other enzymes (Davies and Henrissat, 1995; Lynd et al., 2002). Among them, endoglucanases/cellulases are of particular significance because their action leads to a considerable reduction in the degree of cellulose polymerization, exponentially increasing the concentration of substrates for other enzymes (Béguin and Aubert, 1994). According to the Carbohydrate-Active Enzymes Database (CAZy[1]), among 165 GH families (as of October 2019), endoglucanases are members of the GH5–9, GH12, GH44, GH45, GH48, GH51, GH74, and GH124 families (Lombard et al., 2014).

Archaea are still in the shadow of *Bacteria* in terms of public opinion, the depth of scientific knowledge, and their impact on industrial biotechnology (Straub et al., 2018). There are several reasons for this, among which are the historical development of science, difficulties in isolation and cultivation because the majority of cultivated archaea are still extremophiles and the lack of facile archaeal genetic tools (Straub et al., 2018). Thus, despite the fact that archaea are similar to bacteria in diversity and abundance (Offre et al., 2013), their metabolic variability and industrial potential are lower than those of bacteria. In particular, despite the fact that many bacterial and eukaryal cellulolytic microorganisms have been discovered to date, our knowledge about the degradation of cellulose by archaea is limited. Utilization of cellulose as a growth substrate was demonstrated for several halophilic and natronophilic

representatives of the *Natrialbales* and *Halobacteriales* orders (Sorokin et al., 2015, 2018, 2019) and for hyperthermophilic *Thermococcales* (Mardanov et al., 2009; Gavrilov et al., 2016). A few of the *Thermococcus* and *Pyrococcus* cellulases belong to GH5 and GH12 (Kim and Ishikawa, 2010; Nakahira et al., 2013; Kataoka and Ishikawa, 2014; Gavrilov et al., 2016). As for the *Crenarchaeota*, our current knowledge is restricted to two publications that report a relatively weak growth on cellulose for representatives of *Desulfurococcales* (Perevalova et al., 2005; Kochetkova et al., 2016). Surprisingly, in the genomes of these archaea, no genes of known cellulase families were found that hide their cellulose hydrolysis mechanisms (Mardanov et al., 2012; Susanti et al., 2012). On the other hand, cellulase genes were found in the genomes of some other crenarchaeota, for which the growth on cellulose was not shown. For example, several GH12 family endoglucanases from *Sulfolobus solfataricus* and *S. shibatae* were heterologously expressed in *Escherichia coli* (Limauro et al., 2001; Girfoglio et al., 2012; Boyce and Walsh, 2018). Also using sequence-based and functional metagenomics approaches, a few of the endoglucanases from the GH12 family as well as a novel GH representative were obtained from hyperthermophilic microbial consortia (Graham et al., 2011; Leis et al., 2015; Suleiman et al., 2019).

Representatives of *Thermofilaceae* are moderately acidophilic and hyperthermophilic crenarchaea are obligately dependent on various components of cells or culture broths of other *Crenarchaeota*, yet they are capable of utilizing polysaccharides as sources of carbon and energy (Zillig et al., 1983; Toshchakov et al., 2015; Kochetkova et al., 2019). Until now, cellulose was not among the substrates utilized by representatives of the family except for the weak growth on Avicel of *Thermofilum* sp. strain 1505 (Kochetkova et al., 2019). In this work, we demonstrate the novel species *Thermofilum adornatum* sp. nov. strain 1910b[T]'s ability to grow on cellulose, including insoluble Avicel and in production cell-bound inducible cellulases. Genomic and proteomic-based surveys revealed four novel cellulases of various families, which were heterologously expressed in *E. coli*, and their cellulolytic activities were confirmed and compared.

MATERIALS AND METHODS

Culture Characterization and Cultivation

Thermofilum strain 1910b[T] was isolated from a Kamchatkan hot spring in 2009 (Dominova et al., 2013). A strictly anaerobic modified freshwater Pfennig basal medium (Podosokorskaya et al., 2011) was used for routine cultivation and all growth experiments. The addition of at least 0.05 g/l of yeast extract and 1/100 (v/v) of culture broth filtrate (CBF) for other crenarchaea such as *Fervidicoccus fontis* (strain 1910a = UNIQEM 1910a), *Pyrobaculum aerophilum* (strain Kam13-1) or *Desulfurococcus amylolyticus* (strain 1221n = DSMZ 18924) was mandatory. The utilization of various organic substrates was tested under standard growth conditions under N₂ in the gas phase (80°C, pH 5.5). Other *Thermofilum* species, such as *T. uzonense* strain 1807-2[T] (Toshchakov et al.,

---

[1]www.cazy.org

2015) and strain 1505 (Kochetkova et al., 2019), were used for comparative growth experiments and were taken from the laboratory collection.

Strain 1910b^T was deposited in DSMZ (German Collection of Microorganisms and Cell Cultures) with the designation DSM 28063, and it was also deposited in JCM (Japanese Collection of Microorganisms) with the designation JCM 19809.

Cloning and Heterologous Expression of the Putative Glycosidase Genes in *Escherichia coli*

Cloning of strain 1910b^T genes *Cel25* (2101 bp), *Cel30* (1120 bp), *Cel40* (1558 bp), and *Cel45* (235 bp) was performed using the aLICator Ligation Independent Cloning (LIC) and Expression System kit (#K1251, Thermo Fisher Scientific). The genes were amplified according to Innis and Gelfand (1990) using the *de novo* designed primers (**Supplementary Table S1**) and strain 1910b^T's genomic DNA as the matrix. For the *Cel40* and *Cel25* genes, a nested PCR (Lualdi and Fasano, 2019) was performed because no PCR product with genomic DNA was obtained. This was possibly due to inconsistent theoretical parameters for the use of the respective primer pairs (low GC% content, T melting, etc.). All primers (except for those that were used for nested PCR) include vector-specific sequences (underlined in the table) complementary to linear vector pLATE 51, which contained N-terminal 6x, His-tag and an enterokinase cleavage site (DDDDKˆ).

The PCR products were purified with the Cleanup Standard Kit (#BC022, Evrogen). The LIC reaction was performed for 5 min at 25°C, upon which the vector with the insert was directly transformed into *E. coli* BL21 (DE3) competent cells. This was done following the instructions of the manufacturer, and a 720 bp control PCR fragment was used to assess the efficiency of the LIC reaction. The presence of cloned genes was confirmed by PCR with DNA from grown *E. coli* colonies as the matrix and specific primers (LIC Forward Sequencing primer 5'TAATACGACTCACTATAGGG and LIC Reverse Sequencing primer 5'GAGCGGATAACAATTTCACACAGG), and this was followed by sequencing (Evrogen).

Recombinant *E. coli* strains containing plasmids with target genes were grown in a Luria-Bertani medium (LB) supplemented with ampicillin (100 µg/ml) and 30 µg/l of chloramphenicol for control. This was induced by adding 1.0 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) to express recombinant proteins to a final concentration of cells corresponding to ∼0.5 at OD 600, and it was then incubated at 25°C for 16 h. A volume of 15 ml of the recombinant cells was harvested by centrifugation at 4,000 *g* for 15 min at 4°C, washed with a 25 mM phosphate buffer (pH 7.5) and resuspended in 0.7 ml of the same buffer with 0.5 M NaCl and 25 mM imidazole. After sonication, the cell extracts were centrifuged (15,000 *g* at 4°C for 15 min), and resulting supernatant, as well as the *E. coli* culture broth, were tested for enzymatic activity. The concentration of the purified protein was determined by the Qubit^{TM} protein Assay Kit (#Q33212, Invitrogen).

Glycosidase Activity Measurements
Native Cellulases
Strain 1910b^T was cultivated at optimal growth conditions with 0.5 g/l of microcrystalline cellulose (Avicel) until it reached a cell density of 5–7*10^6 cells/ml (120 h of incubation). Native glycosidase activities were determined in the culture broth and solution of proteins, which were extracted from cell surfaces (SPF, surface protein fraction) according to Gavrilov et al. (2016). The resulting SPFs were 10-times diluted in 50 mM MES (pH$^{25°C}$ = 5.6) before measurements. The protein concentration was measured using the Qubit^{TM} Protein Assay Kit. Enzyme preparations were mixed with 0.1% (w/v) amorphous cellulose (AMC), prepared according to Sorokin et al. (2015), and 0.1% (w/v) carboxymethyl cellulose (CMC) in 0.05 M MES (pH$^{25°C}$ = 5.6) and incubated at 80°C for 90 h. Aliquots of the reaction mixture were sampled every 20 h and stored at 4°C before activity measurements were taken. Glycosidase activities were determined using a DNS assay (Miller, 1959). D-glucose solutions of various (50–500 µg/ml) concentrations were used to plot a calibration curve. One unit (U) of enzyme activity was defined as the amount of enzyme required to release 1 µmol reducing sugar (glucose) in minute at 80°C under the described conditions. Specific activity was defined as enzyme activity per milligram of protein (U/mg).

Recombinant Cellulases
A qualitative analysis of endoglucanase activities was performed as follows: 50 µl of crude extracts of *E. coli* BL21 (DE3) with recombinant proteins were put into the wells in the agar plates, which contained 2% (w/v) agarose and 0.2% (w/v) CMC as a substrate. The plates were incubated at 80°C for 16 h, and this was followed by staining with 0.2% (w/v) Congo red for 30 min (5 ml for each plate) and destaining with 1 M NaCl three times for 15 min at room temperature. Crude extracts of *E. coli* BL21 (DE3) with empty vectors were used as the control experiment.

Quantitative measurements of cellulolytic and cellobiase activities were performed using the DNS method (see above) and CMC and cellobiose as substrates. The reaction mixtures contained 1980 µl of 0.2% (w/v) substrates in a MES buffer (50 mM, pH 5.6) and 20 µl of recombinant enzymes from crude extracts. The reaction mixtures incubated at the optimal temperature (80°C) for 4 h. Crude extracts of *E. coli* BL21 (DE3) with empty vectors were used as the control experiment.

Thin-layer chromatography (TLC) was used to determine the products of cellulose and its derivative hydrolysis. An incubation of recombinant enzymes, which were prepared with CMC, AMC or cellobiose [1% (w/v) each] as the substrates, was performed the same way as it was for the DNS assay (see above). The products of hydrolysis were separated on an aluminum sheet (20 cm × 20 cm) and 60 silica gel plates (Merck), upon which a solution of butanol, ethanol and H$_2$O (2:2:1, v/v/v) was added. The separation of hydrolysis products was performed in a Latch-lid^{TM} TLC developing chamber. A solution containing glucose, cellobiose, cellotriose, cellotetraose, cellopentaose, and cellohexaose [0.0625% (w/v) each] was used as the marker. After elution, the plates were dried at 42°C for 30 min and then dried again at 65°C for 20 min. Mono- and oligomers

were observed upon spraying the plates with a 0.1% orcinol solution in 5% (v/v) H_2SO_4. This was followed by drying at 75°C for 10–15 min.

Functional Genome Analysis

The genome of strain $1910b^T$ was sequenced in another work (Dominova et al., 2013). The genome was deposited in Genbank under the accession numbers CP006646 and IMG 2554235458. The reconstruction of metabolic pathways was done using the KEGG (Kanehisa et al., 2016) and MetaCyc (Caspi et al., 2014) databases. The predicted protein functions were analyzed using a hidden Markov model-based (HMM-based) homology search via the HMMER web server[2] against the Pfam 27.0 database (Finn et al., 2014) and dbCAN 2.0 web resource (Yin et al., 2012; Lombard et al., 2014) was analyzed against the CAZy database. Transporter predictions were done using a BLAST-search against the transporter classification database TCDB (Saier et al., 2014). A more sensitive search for better function prediction was performed using PSI-BLAST (Altschul et al., 1997), genome context analysis and the presence of specific regulatory sequences. Signal peptides were predicted with the SignalP 4.1 (Petersen et al., 2011) and TatP 1.0 servers (Bendtsen et al., 2005); transmembrane helices were predicted with the HMM-based TMHMM 2.0 web server (Krogh et al., 2001). The HMM-based server that combined transmembrane protein topologies and signal peptide predictors, Phobius (Käll et al., 2004), was also used to verify the respective predictions.

Proteomic Analysis

For proteomic analysis, strain $1910b^T$ was grown under optimal growth conditions (80°C and a pH of 5.7) with 1 g/l of Avicel until it reached a cell density of $5–7*10^6$ cells/ml. The experiment was done in four biological replicates (4 bottles that each contained 1.2 l of culture). The control growth experiment (1 g/l of pyruvate instead of cellulose as a substrate), was made in three replicates. The grown cells were collected by centrifugation at 17,600 g for 20 min and lyzed according to Kulak et al. (2014). The protein concentration in the cell-free culture broth was under the detection limit when measured using a Brandford reagent. Cell proteins (both intracellular and cell-wall) were treated the same way as described by Kulak et al. (2014). NanoLC-MS/MS analysis was performed as described previously (Sidorenko et al., 2018). Raw MS data were analyzed using the MaxQuant software package[3] against the strain $1910b^T$ genomic sequence (CP006646). Following this, proteomic analysis was conducted using R. The script is available in the repository[4].

Intensity-based absolute quantitation (iBAQ) quantitative values of detected proteins were obtained and normalized as summed iBAQ values within one sample, which represent the molar abundance or relative iBAQ (riBAQ) of an identified protein within a sample. For each detected protein, the log2 riBAQ value was calculated (Cijsouw et al., 2018). To determine the similarity of the samples' expression profiles, a hierarchical clustering of the samples based on the correlation coefficients of the riBAQ values was performed. To determine the genes with statistically significant differences in expression, an independent, two-sample t-test with the Benjamini–Hochberg correction was used (Green and Sambrook, 2019).

RESULTS

Cell Morphology and Growth Characteristics

Cells of strain $1910b^T$ were thin, straight filaments 0.15 μm in width and 2–20 μm in length. One sub-polar flagellum was occasionally observed (data not shown). Strain $1910b^T$ was an obligate anaerobe, growing optimally on glucose at a temperature of 80°C and at a pH of 5.5–6.0. The generation time and maximum cell density under optimal growth conditions with glucose as a substrate were ∼5 h and $∼5*10^7$ cells/ml, respectively. The isolate grew on carbohydrates such as starch, β-glucan, Avicel and AMC (**Table 1**) with a final cell yield of $5–15*10^6$ cells/ml (**Figure 1**).

The phylogenetic position of strain $1910b^T$ as well as the average nucleotide identity (ANI) values between $1910b^T$ and validly published *T. pendens* and *T. uzonense* placed strain $1910b^T$ in a separate species of the genus (Kochetkova et al., 2019). Taking into account the phenotypic properties of $1910b^T$ (cellulose, β-glucan, cellobiose, mannose, lactose and pyruvate utilization), we propose a novel species of the genus, *T. adornatum* sp. nov., with the type strain $1910b^T$. A description of the new species is given below.

Native Glycosidase Activity

To reveal its native cellulase activity, *T. adornatum* $1910b^T$ was grown on Avicel for 5 days under optimal growth conditions. Upon centrifugation (17,600 g for 20 min), the cell-free culture broth and cell pellet fractions were incubated with CMC or AMC, and the formation of reduced sugars was measured using the DNS approach. Cellulolytic activity was detected only in the cells with a reducing sugars formation rate of 19.33–49.45 μmol glucose/ml/minute (U) (**Table 2**). The cell wall linked cellulases were washed from the cells' surfaces by treatment with solutions of urea, Triton-X100, Tween-80 and SDS. All except the SDS-washed solutions of surface protein fractions (SPFs) contained active cellulases (**Table 2**).

No cellulolytic activity was detected in either the culture broth or cell pellet fractions of *T. adornatum* $1910b^T$ grown on pyruvate, suggesting that cellulases are cellulose inducible (**Figure 2**).

Genomic Analysis of Sugar Metabolism

The genome of *T. adornatum* $1910b^T$ contained 39 genes, which were predicted to encode CAZymes (Lombard et al., 2014). Seventeen of them encoded putative glycosidases (GH), 16
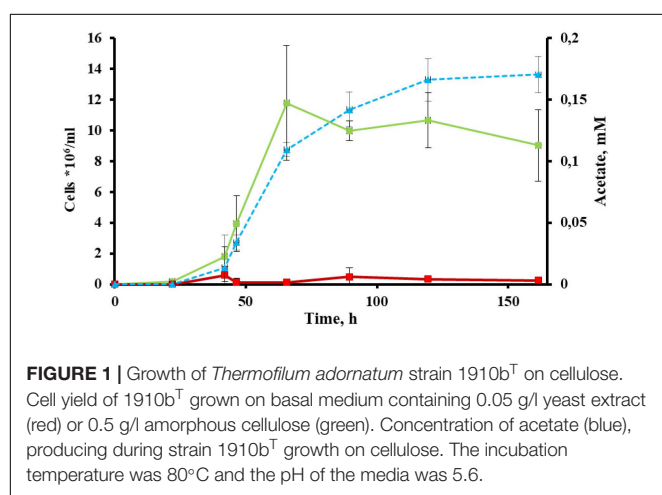

TABLE 1 | Phenotypic characteristics of *Thermofilum* strains.

Characteristics	*T. adornatum* 1910b[T]	*Thermofilum* sp. 1505	*T. pendens* Hvv3[T]	*T. uzonense* 1807-2[T]
Isolation source	Kamchatka	Kamchatka	Iceland	Kamchatka
Min./opt./max. growth temperature (°C)	50/80/95	85	85–90	70/85/90
Min./opt./max. pH for growth	5.3/5.5–6.0/8.5	6.8	4.0/5.0–6.0/6.5	5.5/6.0–6.5/7.0
Growth factors dependence	*Desulfurococcus*, *Pyrobaculum* or *Fervidococcus* CBF, and yeast extract	*Desulfurococcus* CBF, yeast extract, and cysteine	*Thermoproteus tenax* polar lipid fraction and yeast extract	*Desulfurococcus* or *Pyrobaculum* CBF and yeast extract
S^0-dependence	−	−	+	−
Utilized substrates				
AMC	+	−	(+)	−
Avicel	+	+		−*
α-cellulose	−	−		ND
CMC	−	−		−
Filter paper	−	−		−
β-glucan	+	ND	ND	ND
Starch	+	ND	(+g)	+
Mannan	−	ND	ND	−
Glucomannan	−	ND	ND	+
Cellobiose	+	−	ND	−*
Glucose	+	ND	(+g)	+
Sucrose	−	ND	+	ND
Maltose	−	ND	ND	−*
Lactose	+	ND	ND	ND
Mannose	+	ND	ND	−*
Pyruvate	+	ND	ND	ND
G + C mol%	46.5	46.4	57.6	47.9
Genome (Mb)	1.75	1.75	1.78	1.61
Reference	This work	Kochetkova et al., 2019	Zillig et al., 1983; Anderson et al., 2008	Toshchakov et al., 2015

*All strains are strict anaerobes and capable of fermenting yeast extract and peptone. ND, no data available; +, growth; −, no growth (no difference with the control); (+g), the genes, encoding enzymes, presumably involved in degradation of these substrates were found by genome analysis; *, our data.*

FIGURE 1 | Growth of *Thermofilum adornatum* strain 1910b[T] on cellulose. Cell yield of 1910b[T] grown on basal medium containing 0.05 g/l yeast extract (red) or 0.5 g/l amorphous cellulose (green). Concentration of acetate (blue), producing during strain 1910b[T] growth on cellulose. The incubation temperature was 80°C and the pH of the media was 5.6.

encoded glycosyl transferases (GT), 5 encoded carbohydrate esterases (CE), and 1 encoded a protein with auxiliary activities (AA). Four putative GHs, encoded by *N186_RS00270* (GH113), *N186_RS08040* (GH57), *N186_RS08045* (GH13), and *N186_RS08615* (GH16), were predicted to be extracellular proteins. The remaining putative GHs were predicted to be intracellular because no signal peptides were predicted by SignalP, TatP or Phobius (**Supplementary Table S2**). N186_RS08040 with GH57 and N186_RS08045 with GH13 domains were probably involved in starch degradation. Their genes were located in the cluster *N186_RS08010-8060*, also encoded various subunits of ABC transporters, possibly involved in the import of maltooligosaccharides. Another putative GHs of the GH57 family (*N186_RS01555* and *N186_RS01850*) probably participated in the intracellular hydrolysis of maltooligosaccharide or glycogen. Three proteins that presumably participated in both alpha-mannan degradation and utilization were identified. The genes of two of them (*N186_RS07230* and *N186_RS07265*) were co-located with the sugar-specific transcriptional regulator *TrmB* (*N186_RS07255*), while the gene of putative alpha-mannosidase (*N186_RS07400*), which contained two GH38 domains, was clustered with the gene of hypothetical fructose-bisphosphate aldolase (*N186_RS07295*). The close location of the carbohydrate ABC transporter and sugar permease

TABLE 2 | Native cellulases activity of *T. adornatum* 1910b[T].

Fraction	Enzyme activity, U	
	AMC	CMC
Cells	19.33	49.45
Supernatant	0	0
SPF (SDS)	0	0
SPF (Triton-X100)	18.7	20.4
SPF (Urea)	21.39	37
SPF (Tween-80)	8.16	9

(*N186_RS07240*, *N186_RS07245* and *N186_RS07250*) genes suggest the co-action of all these proteins in alpha-mannan degradation. The following proteins were predicted to participate in beta-linked sugar hydrolysis including cellulose and its derivatives: a GH1 glycosidase (*N186_RS00340*, *Cel40*) which gene was clustered with ABC transporters and permease genes *N186_RS00315*, *N186_RS00320*, *N186_RS00325*, and *N186_RS00330*. Another GH1 (*N186_RS06555*, putative beta-galactosidase) gene was co-located with the genes, encoding a PTS transporter (*N186_RS06510*), a sugar-phosphate permease (*N186_RS06540*), and a galactokinase (*N186_RS06575*). The genes of three hypothetical proteins, one of which was homologous to transporters (*N186_RS00220*), another – sugar-specific transcriptional regulators (*N186_RS00235*), while the third one *N186_RS00230* (*Cel30*) lacks any detectable characterized homologs were co-located with a gene, encoded a putative β-glucosidase (*N186_RS00225*, *Cel25*) of the GH3 family. A gene of putative endo-1-3-β-glucanase (*N186_RS08615*) containing a GH16 domain was located near two genes, encoding hypothetical protein N186_RS08625 and DUF86-containing N186_RS08645. No genes that encoded GHs of families contained biochemically characterized cellulases were found in the 1910b[T] genome.

The central carbohydrate metabolism of *T. adornatum* 1910b[T] was similar to other hyperthermophilic archaea (Bräsen et al., 2014). Glucose utilization occurred via the archaeal type Embden–Meyerhof–Parnas (Ahmed et al., 2005), for which all necessary genes were found. The first reaction of this pathway involves hexokinases of the ROK family N186_RS05865 which phosphorylates hexoses such glucose and mannose. Isomerization of glucose-6-phosphate into fructose-6-phosphate was catalyzed by phosphoglucose/phosphomannose-isomerase N186_RS04475, which is capable of using both glucose and mannose as substrates. Phosphofructokinase N186_RS07290 of *T. adornatum* 1910b[T] is probably an ATP-dependent enzyme belonging to the PFK-B family, which representatives are known to be distributed among *Crenarchaeota* (Bräsen et al., 2014). *Thermofilum*'s fructose-1,6-bisphosphate aldolase N186_RS07295 belongs to the family of the archaeal type class I aldolases. The oxidation of glyceraldehyde-3-phosphate occurred in a one-step, irreversible reaction catalyzed by GAPOR (ferredoxin-dependent glyceraldehyde-3-phosphate oxidoreductase N186_RS1755, N186_RS8210, N186_RS1345, and N186_RS0640).
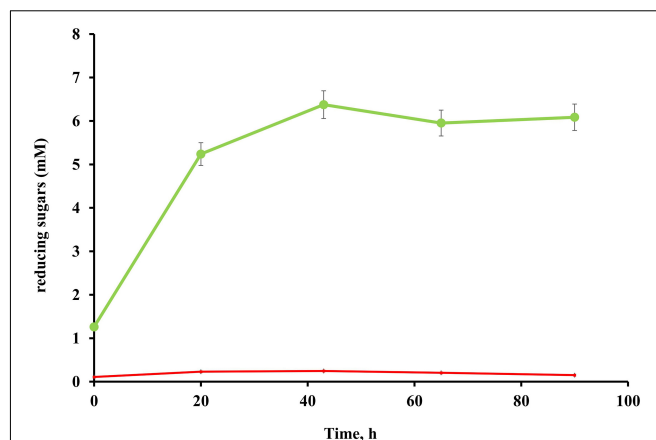
FIGURE 2 | Endoglucanase activity of *T. adornatum* 1910b[T], grown on Avicel (green), or pyruvate (red). The protein concentrations were normalized between the experiments. In both cases the strain was grown under optimal growth conditions. For enzymatic activity measurements CMC was used as the substrate, the incubation time was 90 h, temperature and pH were 80°C and pH 5.6, respectively.

Proteomic Analysis

Despite the fact that *T. adornatum* 1910b[T] grew on cellulose as a sole energy and carbon source for synthesizing cell-bound cellulases, no currently known cellulase-encoding genes were found in its genome. To reveal the proteins involved in cellulose degradation and cellulases in particular, strain 1910b[T] was grown on Avicel or pyruvate (as the control), followed by LC-MS/MS proteomics. Analysis of experimental and control proteomes revealed 139 of 1215 genes, the expression of which was higher on cellulose than on pyruvate (109 genes were downregulated). Most of the upregulated genes (84 of 139) encoded hypothetical proteins, ABC transporters, permeases, and other transporters and proteins involved in ribosome and nucleotide biosynthesis. Thirteen, eight and four proteins encoded by upregulated genes were predicted to be secreted by Phobius, TatP, and SignalP, respectively (**Supplementary Table S3**). The majority (12 of 17) of glycosidase genes were non-regulated or even downregulated on cellulose (**Supplementary Table S4**). Among upregulated genes, three (*N186_RS08540*, *N186_RS08045* and *N186_RS01850*) were predicted to encode enzymes, capable of hydrolyzing maltose, starch and its derivatives; two (*N186_RS07750* and *N186_RS04390*) were glycosyl transferases with a predicted inverting mechanism (thus, they were irreversible); and only one protein (N186_RS00340, Cel40) was annotated as a representative of the GH1 family, which means that it was possibly involved in beta-glucans hydrolysis.

An in-depth genomic analysis of these upregulated genes and their genomic context revealed four candidates [*Cel25*, *Cel30*, *Cel40*, and *N186_RS00345* (*Cel45*)] (**Table 3**), which were the most likely to be involved in cellulose hydrolysis. The selection of putative cellulases was based on a combination of the following criteria: (1) a high ratio of experimental (cellulose) riBAQ values to the control (pyruvate) riBAQ

TABLE 3 | Candidate proteins, proposed to be involved in cellulose hydrolysis by *T. adornatum* 1910b^T.

Gene ID	riBAQ_cel/riBAQ_contr	Function	Pfam	Localization	Gene length, bp
N186_RS00340 (Cel40)	44.8	Glycoside hydrolase family 1 protein	GH1	Intracellular	1558
N186_RS00345 (Cel45)	11.3	Hypothetical protein	No hits	Intracellular	235
N186_RS00230 (Cel30)	3.2	Hypothetical protein	No hits	Extracellular	1120
N186_RS00225 (Cel25)	1.0	Glycoside hydrolase	GH3	Intracellular	2101

values (riBAQ_cel./riBAQ_contr., **Figure 3**), (2) the presence of detectable GH domains of families with known activity against β-linked carbohydrates, (3) extracellular localization (predictions based on signal peptide and N-terminal transmembrane helix) and (4) the gene neighborhood with genes that encoded proteins involved in carbohydrate metabolism. It should be noted that the chosen genes should not necessarily need to fit all of the criteria.

## Glycosidase Activity of the Recombinant Proteins of *Thermofilum adornatum* 1910b^T

The four selected genes were heterologously expressed in *E. coli* BL21 (DE3) using the pLATE expression system (pLATE51 with N-terminal His$_{6x}$-tag). After the destruction of *E. coli* cells and centrifugation of cell debris, only Cel45 was active on the agar plate with CMC as a substrate (**Supplementary Figure S1**).

In contrast, a quantitative activity measurement using a DNS assay with CMC as a substrate revealed the endoglucanase activity of the Cel40, Cel45, and Cel30 proteins (**Figure 4A**).

Moreover, the most active enzyme was Cel40 (37.36 U/mg). The same experimental setup using cellobiose instead of CMC as a substrate (**Figure 4B**) showed three proteins (Cel40, Cel30 and Cel25) that possessed beta-glucosidase activity (15.26, 6.10, and 4.17 U/mg, respectively).

Thin-Layer Chromatography Analysis of Cellulose and Cellobiose Hydrolysis Products
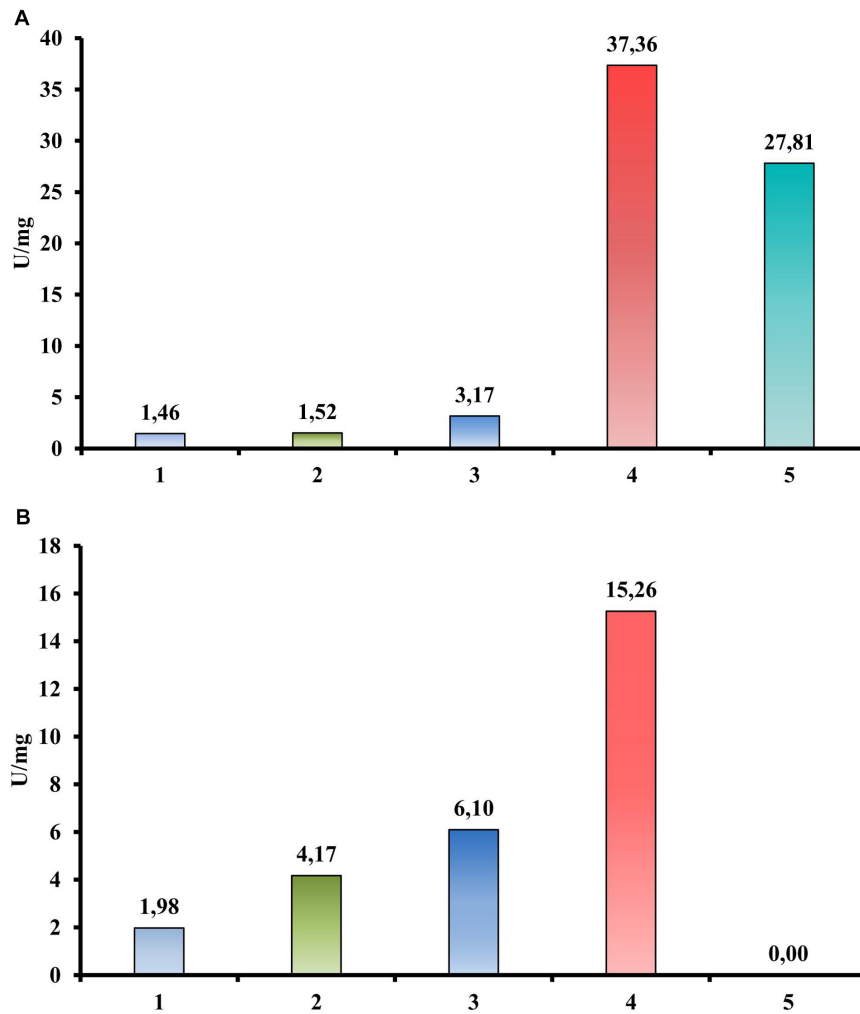
The TLC of the products of CMC hydrolysis by the four studied proteins revealed the production of glucose (C1), cellobiose (C2), and cellotriose (C3) by Cel40 and the production of C2–C3 by Cel25 and Cel30. In its turn, Cel45 produced no detectable C1–C6 products (**Supplementary Figure S2**). Analysis of the AMC hydrolysis products revealed a formation of C1 and C2 due to the action of the Cel40, no detectable C1-C6 sugars were observed after incubation of AMC with three other enzymes. Hydrolysis of cellobiose was observed only for Cel40 and Cel25 (**Supplementary Figure S2**).
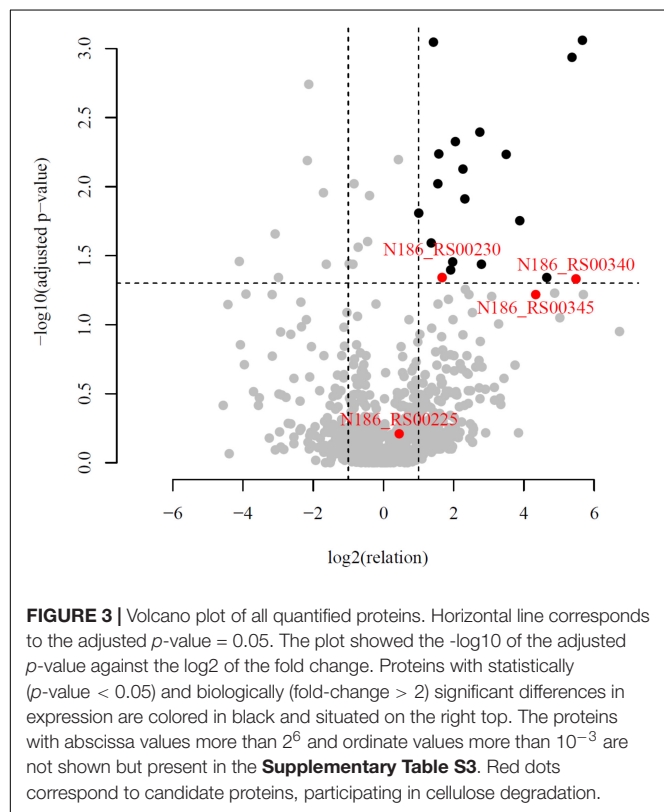
DISCUSSION

The most extremophilic and thermophilic microorganisms are represented by archaea. Despite high demand, linked with their extreme resistance to harsh conditions, archaea still cannot be regarded as key players in bioindustry for several reasons (see section "Introduction"). Yet some progress has been made recently, and this progress is mainly linked with the development of novel archaeal genetic tools (Straub et al., 2018) and the progress in isolation of novel archaea. The latter is one of the crucial (Straub et al., 2018) yet difficult steps toward the involvement of archaea in bioindustry. An isolation of novel extremophilic archaea, especially those of deep phylogenetic lineages and/or possessing rare or novel properties, gives us a good opportunity to develop novel technologies or improve current ones. The crenarchaeal family *Thermofilaceae* is a deeply branching lineage within the order *Thermoproteales*, which consists so far of only one genus (*Thermofilum*). This genus has only two validly published species, *T. pendens* and *T. uzonense*, which were isolated from hot springs in Iceland (Zillig et al., 1983) and Kamchatka (Toshchakov et al., 2015). Representatives of *Thermofilaceae* were thought to be commensals, growing on peptides and simple sugars and requiring undetermined growth factors provided by other crenarchaea (*Thermoproteus*, *Desulforococcus*, *Fervidococcus*, and *Pyrobaculum*, in particular) because the genes of many protein contribute to the biosynthetic

FIGURE 3 | Volcano plot of all quantified proteins. Horizontal line corresponds to the adjusted *p*-value = 0.05. The plot showed the -log10 of the adjusted *p*-value against the log2 of the fold change. Proteins with statistically (*p*-value < 0.05) and biologically (fold-change > 2) significant differences in expression are colored in black and situated on the right top. The proteins with abscissa values more than 2^6 and ordinate values more than 10^{-3} are not shown but present in the **Supplementary Table S3**. Red dots correspond to candidate proteins, participating in cellulose degradation.

FIGURE 4 | Hydrolytic activities of recombinant proteins from *T. adornatum* 1910b^T, measured by DNS. **(A)** CMC is a substrate. **(B)** Cellobiose is a substrate. 1, crude extract of *Escherichia coli* cells with empty vector (control); 2, crude extract of *E. coli*, containing Cel25; 3, crude extract of *E. coli*, containing Cel30; 4, crude extract of *E. coli*, containing Cel40; 5, crude extract of *E. coli*, containing Cel45. Incubation: 3 h at 80°C.

pathways for purines, amino acids, and cofactors were absent in their genomes (Anderson et al., 2008). On the other hand, it was shown that the *Thermofilum* species can grow lithotrophically, gaining energy from carbon monoxide (Kochetkova et al., 2019). Moreover, Toshchakov et al. (2015) revealed that *T. uzonense* is capable of growing on two polysaccharides: starch and glucomannan. The growth experiments correlated with *T. uzonense*'s genomic analysis, during which the genes encoding various GHs, including amylases and mannanases, were found. Moreover, in genomes of *T. pendens* (Anderson et al., 2008) and *T. uzonense* (Toshchakov et al., 2015), determinants of cellulose degradation were observed, but no evidence of positive growth on cellulose was shown. *Thermofilum* sp. strain 1505 grew on Avicel (Kochetkova et al., 2019), but its growth was weak, and no additional evidence was obtained to support it. The novel species *T. adornatum* 1910b^T, which is described here, is able to utilize various carbohydrates, including polysaccharides

such as starch, beta-glucan and cellulose (both amorphous and microcrystalline). It is worth mentioning that the final cell yields of the three crenarchaeal species previously reported as capable of growing on cellulose, *Desulforococcus fermentans*, *Thermogladius calderae* and *Thermofilum* sp. strain 1505, were only 1.5–3 times higher than in the control experiment (the same medium without cellulose) (Perevalova et al., 2005; Kochetkova et al., 2016, 2019). This may suggest that the growth on cellulose was determined by the presence of beta-glucosidases, which acted non-specifically, or that the optimal growth conditions were not ascertained. By contrast, the growth of strain 1910b^T on cellulose was 10 times higher when compared to a basal medium without cellulose (**Figure 1**), which cannot be explained by the non-specific action of other enzymes. Moreover, in the case of 1910b^T, these growth experiments were supported by several activity-based approaches as well as genomic and proteomic analyses.

Native enzyme activities were measured in *T. adornatum* 1910b^T cells, which were grown on Avicel using the DNS approach. No enzyme activities were detected in the cell-free culture broth, which implies the enzymes were bound to the cell's envelopes via a lipid anchor or a C-terminal transmembrane region or by other mechanisms (Szabo and Pohlschroder, 2012). Cellulases, anchored to the cells and the same time bound to its substrate give additional advantage to cellulolytic microorganism since it will be the first to consume the hyrolysis products (Lu et al., 2006). The active cellulases were washed using various buffers and chaotropic agents, and the highest reduced sugar releasing rates were associated with surface proteins that were solubilized by urea or Triton X 100 (**Table 2**). In comparison to AMC, a sodium salt of CMC resulted in a higher rate of reducing sugars formation that can be explained by the higher solubility and accessibility of the latter. An absence of cellulase activities in the cells grown on pyruvate indicates that the cellulases of strain 1910b^T are cellulose induced.

Despite significant deviation in the spectra of polysaccharides utilized by all studied *Thermofilum* strains (**Table 1**), the number and variation of genomes in the CAZymes genes are quite similar (**Supplementary Figure S3**). Thus, the variety of utilized substrates might be determined by yet uncharacterized enzymes, and the overall hydrolytic potential of the genus is not completely recognized. This proposition is consistent with about 30% of protein-coding genes without function prediction in all four *Thermofilum* representatives, according to the Integrated microbial genomes (IMG) portal (**Supplementary Figure S4**). All *Thermofilum*'s genomes encode the GHs of families GH1, GH3, GH4, GH38, GH57, GH65, and GH109. None of these families include currently known cellulases, moreover taking into account that, except for 1910b^T (and possibly strain 1505), none of the strains are capable of growing on cellulose, one would not expect the representatives of these GH families to play a significant role in cellulose hydrolysis in 1910b^T. Despite the fact that both the *T. adornatum* 1910b^T and the *T. uzonense* 1807-2^T genomes contain the genes of putative mannanases of the GH113 and GH130 families (*N186_RS00270* and *N186_RS07230* in 1910b^T and *MA03_02300* and *MA03_02580* in the 1807-2 genomes, respectively), only *T. uzonense* can grow on glucomannan (Toshchakov et al., 2015); thus, their function in 1910b^T is unclear. Still, it is unlikely that these enzymes are involved in cellulose hydrolysis. To our surprise, we were unable to find homologs of a putative cellulase (GH12 family) gene previously found in the *T. pendens* genome (Anderson et al., 2008) in *T. adornatum* 1910b^T genome (**Supplementary Figure S3**). It is possible that this enzyme has a different function because *T. pendens* Hrk5^T was not shown to grow on cellulose. Thus, despite the fact that *T. adornatum* 1910b^T grows on cellulose, producing cell-bound cellulases, neither genes that encode known cellulases nor GHs distantly related to beta-acting enzymes and exclusively present in *T. adornatum* 1910b^T were found in its genome. This led to the proposal that unknown cellulose-degrading enzymes were involved. A more detailed analysis of genomic data as well as an evaluation of genes, which are upregulated during growth on Avicel using proteomics, revealed four candidate endoglucanases (**Table 3**).

Three of them had higher cellulose/control riBAQ ratios than the threshold. Despite the fact that the cellulose/control riBAQ ration for Cel25 protein was lower than the threshold, it had a detectable GH domain and its gene was co-located with another candidate, *Cel30*, which encoded the only extracellular enzyme among the four candidates. In practice, one should be aware that protein localization prediction servers can still be highly inaccurate with archaeal sequences due to the limited amount of experimentally verified data (Bagos et al., 2009). Two of candidate proteins (Cel25 and Cel40) possessed GH1 and GH3 domains, both families contain enzymes with exo-glucanase activities, which may be involved in cellulose degradation (Varghese et al., 1999; Yernool et al., 2000). Two other proteins (Cel45 and Cel30) had no detectable homologs with predicted functions. The proteins Cel40 and Cel45, which were predicted to be intracellular but had high cellulose/control riBAQ ratios, and their genes were co-located. The four candidate cellulase genes were heterologously expressed in *E. coli* BL21 (DE3), and the respective activities of the recombinant proteins were measured. Both DNS activity measurements and TLC of the products of CMC, AMC, and cellobiose hydrolysis suggest that all four recombinant enzymes participated in cellulose degradation. Among them, the Cel25 and Cel30 proteins were likely relatively low-activity exoglucanases with even lower side activities (endoglucanase and beta-glucosidase). Due to its high activity on CMC as well as inability of hydrolizing or producing cellobiose, the Cel45 protein was an endoglucanase. Surprisingly, Cel40 had the highest activity against both CMC and cellobiose, indicating its lesser dependence on the substrate's length. This observation is in accordance with the most variable set of detectable cellooligosaccharides (**Supplementary Figure S2**) produced during the action of Cel40. It is probable that both enzymes, Cel40 and Cel45, act synergistically and that their genes are coherently regulated by the presence of a substrate. The synergistic action of the GHs of various substrate specificities and affinities is well known for multienzyme complexes, as cellulosomes (Bayer et al., 2008) or SUS (Shipman et al., 2000), as well as for single polypeptides containing several domains of various functions (e.g., Gavrilov et al., 2016), which degrade huge, rigid or complex polysaccharides. Taking into account that (1) recombinant Cel25 and Cel30 were much less active than Cel40 and Cel45; (2) their gene expressions were (almost) not induced by the presence of CMC, according to proteomic analysis; and (3) the cellulolytic activity of *T. adornatum* 1910b^T's SPFs was inducible, we suggest that Cel40 and Cel45 play a major role in cellulose degradation, especially in crucial steps of intramolecular hydrolysis. We suggest that Cel25 and Cel30, on the other hand, are only accessory enzymes. Further work is needed to verify this.

An analysis of these 4 enzymes gene distribution within the *Thermofilum* genomes revealed that *Cel25* and *Cel40* were present in all of them, and they were inherited vertically because the pairwise identities of *Cel25* and *Cel40* were similar to the Average Amino-acid Identity values (**Supplementary Figure S5**). In their turn, *Cel30* and *Cel45* were present only in the *T. adornatum* strain 1910b^T and strain 1505 genomes (**Supplementary Table S5**). From one standpoint, this observation strongly supports the major role of *Cel45* in

cellulose hydrolysis. From another, this observation makes it unclear why strain 1505 only weakly grew on Avicel. The most probable answers have to do with the differences in regulation and/or transport of cellooligosaccharides between these two strains. The incompleteness of the 1505 growth experiments cannot be disregarded.

Thus, *Thermofilum adornatum* 1910b^T is the first *Thermofilum* representative capable of cellulose degradation. This was determined by a set of four GHs, two of which represented GH families with previously unknown cellulase activities. The other two were known as hypothetical proteins. The enzymes seem to have different roles in cellulose hydrolysis by strain 1910b^T due to differences in substrate specificities as well as various potential mechanisms of secretion, which can affect their localizations and thus their modes of introduction to their substrates.

Description of *Thermofilum adornatum* sp. nov.

Thermofilum adornatum (a.dor.na'tum N.L. neut. adj. *adornatum*, often adorned with clubs at the ends of its cells).

Cells are thin filaments that are 0.15 μm in width and 2–20 μm in length with one flagellum. It is a strict anaerobe. Its temperature, pH and NaCl ranges for growth are 50–95°C, 5.3–8.5 and 0–2.5%, respectively. Its optimal growth conditions include a temperature of 80°C, a pH of 5.5–6.0 and an absence of NaCl. It grows chemoorganoheterotrophically on peptone, yeast extract, AMC, microcrystalline cellulose (Avicel), β-glucan, starch, cellobiose, glucose, lactose, mannose and pyruvate. It does not utilize tryptone, casein, α-cellulose, CMC, lichenan, gelatin, chitin, chitosan, xylan, keratin, mannan, glucomannan, glycerol, sucrose, maltose, xylose, and arabinose. Yeasts extract and CBF of *Desulfurococcus*, *Pyrobaculum* or *Fervidococcus* are required for its growth. The type strain is 1910b^T (= DSM 28063^T = JCM 19809^T), which was isolated from a Kamchatkan hot spring (Russia). The genome size is 1.75 Mb. The G + C content of its DNA is 46.5 mol. %. The genome sequence of the strain is deposited in GenBank and IMG under the accession numbers CP006646 and 2554235458, respectively.

REFERENCES

Ahmed, H., Ettema, T. J. G., Tjaden, B., Geerling, A. C. M., Van Der Oost, J., and Siebers, B. (2005). The semi-phosphorylative entner-doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem. J.* 390, 529–540. doi: 10.1042/BJ20041711

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Anderson, I., Rodriguez, J., Susanti, D., Porat, I., Reich, C., Ulrich, L. E., et al. (2008). Genome sequence of *Thermofilum pendens* reveals an exceptional loss of biosynthetic pathways without genome reduction. *J. Bacteriol.* 190, 2957–2965. doi: 10.1128/JB.01949-07

Bagos, P. G., Tsirigos, K. D., Plessas, S. K., Liakopoulos, T. D., and Hamodrakas, S. J. (2009). Prediction of signal peptides in archaea. *Protein Eng. Design Select.* 22, 27–35. doi: 10.1093/protein/gzn064

Bayer, E. A., Lamed, R., White, B. A., and Flints, H. J. (2008). From cellulosomes to cellulosomics. *Chem. Rec.* 8, 364–377. doi: 10.1002/tcr.20160

DATA AVAILABILITY STATEMENT

AUTHOR CONTRIBUTIONS

KZ, TK, and IK conceived the study. KZ and TK performed the microbiological and biochemical experiments. KZ, UP, and IK did bioinformatics analyses. RZ, KZ, and UP performed the proteomics. OP isolated the archaeon into a pure culture. All authors were involved in writing and reviewing the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02972/full#supplementary-material

Béguin, P., and Aubert, J. P. (1994). The biological degradation of cellulose. *FEMS Microbiol. Rev.* 13, 25–58. doi: 10.1016/0168-6445(94)90099-X

Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005). Prediction of twin-arginine signal peptides. *BMC Bioinform.* 6:167. doi: 10.1186/1471-2105-6-167

Boyce, A., and Walsh, G. (2018). Expression and characterisation of a thermophilic endo-1,4-β-glucanase from *Sulfolobus shibatae* of potential industrial application. *Mol. Biol. Rep.* 45, 2201–2211. doi: 10.1007/s11033-018-4381-7

Bräsen, C., Esser, D., Rauch, B., and Siebers, B. (2014). Carbohydrate metabolism in archaea: current insights into unusual enzymes and pathways and their regulation. *JAMA Ophthalmol.* 132, 326–331. doi: 10.1128/MMBR.00041-13

Cabrera, M. Á, and Blamey, J. M. (2018). Biotechnological applications of archaeal enzymes from extreme environments. *Biol. Res.* 51:37. doi: 10.1186/s40659-018-0186-3

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and

the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 42, D742–D753. doi: 10.1093/nar/gkt1103

Cijsouw, T., Ramsey, A. M., Lam, T. K. T., Carbone, B. E., Blanpied, T. A., and Biederer, T. (2018). Mapping the proteome of the synaptic cleft through proximity labeling reveals new cleft proteins. *Proteomes* 6:E48. doi: 10.3390/proteomes6040048

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853–859. doi: 10.1016/S0969-2126(01)00220-9

Dominova, I. N., Kublanov, I. V., Podosokorskaya, O. A., Derbikova, K. S., Patrushev, M. V., and Toshchakov, S. V. (2013). Complete genomic sequence of "*Thermofilum adornatus*" strain 1910bT, a hyperthermophilic anaerobic organotrophic crenarchaeon. *Genome Announc.* 1:e0726-13. doi: 10.1128/genomeA.00726-13

Egorova, K., and Antranikian, G. (2005). Industrial relevance of thermophilic Archaea. *Curr. Opin. Microbiol.* 8, 649–655. doi: 10.1016/j.mib.2005.10.015

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223

Gavrilov, S. N., Stracke, C., Jensen, K., Menzel, P., Kallnik, V., Slesarev, A., et al. (2016). Isolation and characterization of the first xylanolytic hyperthermophilic euryarchaeon *Thermococcus* sp. strain 2319x1 and its unusual multidomain glycosidase. *Front. Microbiol.* 7:552. doi: 10.3389/fmicb.2016.00552

Girfoglio, M., Rossi, M., and Cannio, R. (2012). Cellulose degradation by *Sulfolobus solfataricus* requires a cell-anchored endo-β-1-4-glucanase. *J. Bacteriol.* 194, 5091–5100. doi: 10.1128/JB.00672-12

Graham, J. E., Clark, M. E., Nadler, D. C., Huffer, S., Chokhawala, H. A., Rowland, S. E., et al. (2011). Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat. Commun.* 2:375. doi: 10.1038/ncomms1373

Green, M. R., and Sambrook, J. (2019). Nested polymerase chain reaction (PCR). *Cold Spring Harb. Protoc.* 2019, 175–178. doi: 10.1101/pdb.prot095182

Haki, G. D., and Rakshit, S. K. (2003). Developments in industrially important thermostable enzymes: a review. *Bioresour. Technol.* 89, 17–34. doi: 10.1016/S0960-8524(03)00033-6

Innis, M. A., and Gelfand, D. H. (1990). "Optimization of PCRs," in *PCR Protocols: A Guide to Methods and Applications*, eds M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, (London: Academic Press), 3–12. doi: 10.1016/b978-0-12-372180-8.50005-6

Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* 338, 1027–1036. doi: 10.1016/j.jmb.2004.03.016

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Kang, Q., Appels, L., Tan, T., and Dewil, R. (2014). Bioethanol from lignocellulosic biomass: current findings determine research priorities. *Sci. World J.* 2014:298153. doi: 10.1155/2014/298153

Kataoka, M., and Ishikawa, K. (2014). Complete saccharification of β-glucan using hyperthermophilic endocellulase and β-glucosidase from *Pyrococcus furiosus*. *Biosci. Biotechnol. Biochem.* 78, 1537–1541. doi: 10.1080/09168451.2014.923300

Kim, H. W., and Ishikawa, K. (2010). Complete saccharification of cellulose at high temperature using endocellulase and β-glucosidase from *Pyrococcus* sp. *J. Microbiol. Biotechnol.* 20, 889–892. doi: 10.4014/jmb.0912.12020

Kochetkova, T. V., Kublanov, I. V., Toshchakov, S. V., Osburn, M. R., Novikov, A. A., Bonch-Osmolovskaya, E. A., et al. (2016). Thermogladius calderae gen. nov., sp. nov., an anaerobic, hyperthermophilic crenarchaeote from a Kamchatka hot spring. *Int. J. Syst. Evol. Microbiol.* 66, 1407–1412. doi: 10.1099/ijsem.0.000916

Kochetkova, T. V., Mardanov, A. V., Sokolova, T. G., Bonch-Osmolovskaya, E. A., Kublanov, I. V., Kevbrin, V. V., et al. (2019). The first crenarchaeon capable of growing by anaerobic carbon monoxide oxidation coupled with H2 production. *Syst. Appl. Microbiol.* [in press].

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* 3, 319–324. doi: 10.1038/nmeth.2834

Leis, B., Heinze, S., Angelov, A., Trang Pham, V. T., Thürmer, A., Jebbar, M., et al. (2015). Functional screening of hydrolytic activities reveals an extremely thermostable cellulase from a deep-sea archaeon. *Front. Bioeng. Biotechnol.* 3:95. doi: 10.3389/fbioe.2015.00095

Limauro, D., Cannio, R., Fiorentino, G., Rossi, M., and Bartolucci, S. (2001). Identification and molecular characterization of an endoglucanase gene, celS, from the extremely thermophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* 5, 213–219. doi: 10.1007/s007920100200

Lombard, V., Golaconda, R. H., Drula, E., Coutinho, P., and Henrissat, B. (2014). BT2824 - the carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Lu, Y., Zhang, Y.-H. P., and Lynd, L. R. (2006). Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16165–16169. doi: 10.1073/pnas.0605381103

Lualdi, M., and Fasano, M. (2019). Statistical analysis of proteomics data: a review on feature selection. *J. Proteom.* 198, 18–26. doi: 10.1016/j.jprot.2018.12.004

Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* 66, 739–739. doi: 10.1128/mmbr.66.4.739.2002

Maleki, S. S., Mohammadi, K., and Ji, K. S. (2016). Characterization of cellulose synthesis in plant cells. *Sci. World J.* 2016:8641373. doi: 10.1155/2016/8641373

Mardanov, A. V., Kochetkova, T. V., Beletsky, A. V., Bonch-Osmolovskaya, E. A., Ravin, N. V., and Skryabin, K. G. (2012). Complete genome sequence of the hyperthermophilic cellulolytic crenarchaeon "*Thermogladius cellulolyticus*" 1633. *J. Bacteriol.* 194, 4446–4447. doi: 10.1128/JB.00894-12

Mardanov, A. V., Ravin, N. V., Svetlitchnyi, V. A., Beletsky, A. V., Miroshnichenko, M. L., Bonch-Osmolovskaya, E. A., et al. (2009). Metabolic versatility and indigenous origin of the archaeon *Thermococcus sibiricus*, isolated from a siberian oil reservoir, as revealed by genome analysis. *Appl. Environ. Microbiol.* 75, 4580–4588. doi: 10.1128/AEM.00718-09

Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* 31, 426–428. doi: 10.1021/ac60147a030

Nakahira, Y., Ishikawa, K., Tanaka, K., Tozawa, Y., and Shiina, T. (2013). Overproduction of hyperthermostable β-1,4-endoglucanase from the archaeon *Pyrococcus horikoshii* by tobacco chloroplast engineering. *Biosci. Biotechnol. Biochem.* 77, 2140–2143. doi: 10.1271/bbb.130413

Offre, P., Spang, A., and Schleper, C. (2013). Archaea in biogeochemical cycles. *Ann. Rev. Microbiol.* 67, 437–457. doi: 10.1146/annurev-micro-092412-155614

Peacock, J. P., Cole, J. K., Murugapiran, S. K., Dodsworth, J. A., Fisher, J. C., Moser, D. P., et al. (2013). Pyrosequencing reveals high-temperature cellulolytic microbial consortia in Great Boiling Spring after in situ lignocellulose enrichment. *PLoS One* 8:e59927. doi: 10.1371/journal.pone.0059927

Perevalova, A. A., Svetlichny, V. A., Kublanov, I. V., Chernyh, N. A., Kostrikina, N. A., Tourova, T. P., et al. (2005). *Desulfurococcus fermentans* sp. nov., a novel hyperthermophilic archaeon from a Kamchatka hot spring, and emended description of the genus *Desulfurococcus*. *Int. J. Syst. Evol. Microbiol.* 55, 995–999. doi: 10.1099/ijs.0.63378-0

Petersen, T. N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Podosokorskaya, O. A., Merkel, Y. A., Kolganova, T. V., Chernyh, N. A., Miroshnichenko, M. L., Bonch-Osmolovskaya, E. A., et al. (2011). *Fervidobacterium riparium* sp. nov., a thermophilic anaerobic cellulolytic bacterium isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* 61, 2697–2701. doi: 10.1099/ijs.0.026070-0

Robak, K., and Balcerek, M. (2018). Review of second generation bioethanol production from residual biomass. *Food Technol. Biotechnol.* 56, 174–187. doi: 10.17113/ftb.56.02.18.5428

Robb, F. T., Antranikian, G., Grogan, D. W., and Driessen, A. J. M. (2007). *Thermophiles: Biology and Technology at High Temperatures*. Boca Raton, FL: CRC Press.

Saier, M. H., Reddy, V. S., Tamang, D. G., and Västermark, Å (2014). The transporter classification database. *Nucleic Acids Res.* 42, D251–D258. doi: 10.1093/nar/gkt1097

Shipman, J. A., Berleman, J. E., and Salyers, A. A. (2000). Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J. Bacteriol.* 182, 5365–5372. doi: 10.1128/JB.182.19.5365-5372.2000

Sidorenko, S. V., Ziganshin, R. H., Luneva, O. G., Deev, L. I., Alekseeva, N. V., Maksimov, G. V., et al. (2018). Proteomics-based identification of hypoxia-sensitive membrane-bound proteins in rat erythrocytes. *J. Proteom.* 184, 25–33. doi: 10.1016/j.jprot.2018.06.008

Sims, R. E. H., Mabee, W., Saddler, J. N., and Taylor, M. (2010). An overview of second generation biofuel technologies. *Bioresour. Technol.* 101, 1570–1580. doi: 10.1016/j.biortech.2009.11.046

Sorokin, D. Y., Khijniak, T. V., Elcheninov, A. G., Toshchakov, S. V., Kostrikina, N. A., Bale, N. J., et al. (2019). *Halococcoides cellulosivorans* gen. nov., sp. nov., an extremely halophilic cellulose-utilizing haloarchaeon from hypersaline lakes. *Int. J. Syst. Evol. Microbiol.* 69, 1327–1335. doi: 10.1099/ijsem.0.003312

Sorokin, D. Y., Khijniak, T. V., Kostrikina, N. A., Elcheninov, A. G., Toshchakov, S. V., Bale, N. J., et al. (2018). *Natronobiforma cellulositropha* gen. nov., sp. nov., a novel haloalkaliphilic member of the family Natrialbaceae (class *Halobacteria*) from hypersaline alkaline lakes. *Syst. Appl. Microbiol.* 41, 355–362. doi: 10.1016/j.syapm.2018.04.002

Sorokin, D. Y., Toshchakov, S. V., Kolganova, T. V., and Kublanov, I. V. (2015). Halo(natrono)archaea isolated from hypersaline lakes utilize cellulose and chitin as growth substrates. *Front. Microbiol.* 6:942. doi: 10.3389/fmicb.2015.00942

Straub, C. T., Counts, J. A., Nguyen, D. M. N., Wu, C. H., Zeldes, B. M., Crosby, J. R., et al. (2018). Biotechnology of extremely thermophilic archaea. *FEMS Microbiol. Rev.* 42, 543–578. doi: 10.1093/femsre/fuy012

Suleiman, M., Schröder, C., Klippel, B., Schäfers, C., Krüger, A., and Antranikian, G. (2019). Extremely thermoactive archaeal endoglucanase from a shallow marine hydrothermal vent from Vulcano Island. *Appl. Microbiol. Biotechnol.* 103, 1267–1274. doi: 10.1007/s00253-018-9542-z

Susanti, D., Johnson, E. F., Rodriguez, J. R., Anderson, I., Perevalova, A. A., Kyrpides, N., et al. (2012). Complete genome sequence of *Desulfurococcus fermentans*, a hyperthermophilic cellulolytic crenarchaeon isolated from a freshwater hot spring in Kamchatka, Russia. *J. Bacteriol.* 194, 5703–5704. doi: 10.1128/JB.01314-12

Szabo, Z., and Pohlschroder, M. (2012). Diversity and subcellular distribution of archaeal secreted proteins. *Front. Microbiol.* 3:207. doi: 10.3389/fmicb.2012.00207

Toshchakov, S. V., Korzhenkov, A. A., Samarov, N. I., Mazunin, I. O., Mozhey, O. I., Shmyr, I. S., et al. (2015). Complete genome sequence of and proposal of *Thermofilum uzonense* sp. nov. A novel hyperthermophilic crenarchaeon and emended description of the genus *Thermofilum*. *Stand. Genom. Sci.* 10:122. doi: 10.1186/s40793-015-0105-y

Varghese, J. N., Hrmova, M., and Fincher, G. B. (1999). Three-dimensional structure of a barley β-D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure* 7, 179–190. doi: 10.1016/S0969-2126(99)80024-0

Viikari, L., Alapuranen, M., Puranen, T., Vehmaanperä, J., and Siika-Aho, M. (2007). Thermostable enzymes in lignocellulose hydrolysis. *Adv. Biochem. Eng. Biotechnol.* 108, 121–145. doi: 10.1007/10-2007-065

Yernool, D. A., McCarthy, J. K., Eveleigh, D. E., and Bok, J. D. (2000). Cloning and characterization of the glucooligosaccharide catabolic pathway β-glucan glucohydrolase and cellobiose phosphorylase in the marine hyperthermophile *Thermotoga neapolitana*. *J. Bacteriol.* 182, 5172–5179. doi: 10.1128/JB.182.18.5172-5179.2000

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). DbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451. doi: 10.1093/nar/gks479

Zillig, W., Gierl, A., Schreiber, G., Wunderl, S., Janekovic, D., Stetter, K. O., et al. (1983). The archaebacterium *Thermofilum pendens* represents, a novel genus of the thermophilic, anaerobic sulfur respiring *Thermoproteales*. *Syst. Appl. Microbiol.* 4, 79–87. doi: 10.1016/S0723-2020(83)80035-6