



Szent István University

Ph.D. thesis

**Isolation of lignocellulolytic microbes and biotechnological application of
their enzymes**

Ákos Tóth

**Gödöllő
2016**

Ph.D. School

Name: Ph.D. School of Environmental Sciences

Discipline: Environmental Sciences

Leader of School: Csákiné Dr. Erika Michéli, Ph.D.
Professor, head of department
Szent István University
Faculty of Agricultural and Environmental Sciences
Institute of Environmental Sciences
Department of Soil sciences and Agrochemistry

Supervisors: Dr. József Kukolya
Senior Scientist, head of department
Central Environmental and Food Science Research
Institute
Department of Microbiology

Dr. Balázs Kriszt
Associate professor, director of institute, head of
department
Szent István University
Faculty of Agricultural and Environmental Sciences
Institute of Aquaculture and Environmental Safety
Department of Environmental Safety and Ecotoxicology

.....
Approval of School Leader

.....
Approval of Supervisors

Table of contents

1. Background and aims	2
2. Materials and methods.....	3
2.1. Isolation and identification of cellulolytic and hemicellulolytic microorganisms	3
2.1.1. Isolation of microbes	3
2.1.2. Identification of isolates	3
2.2. Microbiological description of the new species, new genus candidate	4
2.2.1. Phenotypic characterization.....	4
2.2.2. Chemotaxonomical characterization	4
2.3. Genome analysis.....	5
2.4. Examination of <i>Thermobifida</i> endomannanases	5
2.4.1. Cloning of the endomannanase genes	6
2.4.2. Protein expression and purification	7
2.4.3. Biochemical characterization of the endomannanases	7
3. Results	8
3.1 Building of compost originated strain collection	8
3.2 Taxonomical investigation of K13 strain	8
3.3 Results of genome analysis	9
3.4. Examination of endomannanases from three different <i>Thermobifida</i> species.....	11
3.5. New scientific results	13
4. Conclusions, proposals	15
6. Publications	19

1. Background and aims

Lignocellulose is one of the main components of plant cell wall, providing abundant renewable raw material for a range of industries. The main barrier of the lignocellulose application is its complex structure and therefore its inefficient processability. Main components of lignocellulose are lignin, cellulose, and hemicelluloses of diverse composition, like xylan and mannan. The most effective aerobic biodegradation of lignocellulosic biomass occurs in composts. The effective conversion of lignocellulose requires a microbial community and their set of enzymes with different activities. During the composting process a microbial community of bacteria and fungi performs the enzymatic decomposition of lignocellulose, this process is aerobic and contain a thermophilic phase. There are a lot of reasons to examine compost samples if we want to study the depolymerisation process of biopolymer content of lignocellulose or we are searching new, carbohydrate active enzymes or microorganisms. At first there is a very diverse microbial community in the compost piles, these microorganisms possess a complete and effective lignocellulose degrading enzyme system. Furthermore, the thermophilic phase is dominated by microorganism that express thermostable enzymes with high thermal optimum. Genome examination of new microbial isolates provides a possibility to find new metabolic pathways and enzymes of biopolymers decomposition. These enzymes have promising properties at the point of industrial applications, since the use of robust and stable enzymes under these conditions can be more profitable. The possible industrial applications of these enzymes are cellulose based biofuel production, feeding, bleaching without chloride in the paper industry, and the lately more and more pronounced prebiotic production.

I set out the following objectives in my dissertation: Isolation and identification of potential lignocellulolytic bacteria from mesophilic and thermophilic regions of compost piles. Characterization and classification of new species candidate isolates. Making of genome projects of lignocellulolytic *Thermobifida* species and the new species candidate strain(s). Cloning and making an overexpression of *man5A* endomannanase genes from different *Thermobifida* species. Characterizing and comparing Man5A recombinant proteins.

2. Materials and methods

I used the kits and materials described in this section according to the manufacturer's instruction unless otherwise indicated. All chemicals herein used were purchased from Sigma-Aldrich Ltd. unless otherwise indicated.

2.1. Isolation and identification of cellulolytic and hemicellulolytic microorganisms

2.1.1. Isolation of microbes

The sampling was performed from 40-80 °C regions of compost piles. A suspension and dilution series were made from the samples. Bacteria were isolated from composts by culturing on agar media that contain different forms of cellulose and hemicellulose (crystalline cellulose - Avicel, carboxymethyl - cellulose, beechwood xylan, and galactomannan-locust bean gum) as sole carbon source.

2.1.2. Identification of isolates

The bacterial biomass was collected by centrifugation when the liquid culture reached the right density, then DNA isolation was performed with the appropriate kit (MoBio UltraClean Microbial DNA Isolation Kit). The result of the isolation was verified by agarose gel electrophoresis. Amplification of the 16S rDNA was performed by optimized PCR. 27f-1492r universal *Bacteria*

primers were used for the reaction. Sequence analysis of the amplified genes (after the steps of purifying the PCR products, the sequencing reaction, and the ethanol precipitation) was performed by capillary gel electrophoresis (Applied Biosystems® ABI 310 DNA Genetic Analyzer, Applied Biosystems, USA). MEGA6 software was used for the evaluation and manual correction of the sequences, and the resulting sequences were compared with EzTaxon database. Phylogenetic trees were reconstructed using MEGA6 with neighbour-joining methods with Kimura's two parameter calculation model.

2.2. Microbiological description of the new species, new genus candidate

The description of new bacterial species was performed with polyphasic approach. Characterization of new isolates was based on DNA techniques, chemotaxonomical and classical microbiological methods.

2.2.1. Phenotypic characterization

The morphological examination was performed with light microscope and electron microscope. The optimal growth pH (pH 4-11), optimal growth temperature (20-50 °C), and salt tolerance (1-5% NaCl) was determined in TSB media. API 50 CHB was used to test the substrate utilization of the isolates.

2.2.2. Chemotaxonomical characterization

The chemotaxonomical characters were determined in the laboratories of DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen). The biomass for the examinations was provided by me. The appropriate protocols are available on the DSMZ's website (<https://www.dsmz.de/services/services-microorganisms/identification.html>). The guanine-cytosine ratio (G+C) was determined *in silico* based on genome project data.

2.3. Genome analysis

The bacterial strains objected for genome projects originated from different sources. Dr. József Kukolya provided me with the *T. fusca* TM51 and *T. cellulositytica* TB100^T strains, *T. halotolerans* JCM16012^T (same as: YIM90462^T) was purchased from official strain collection, and strain K13 was isolated by myself. The described genome projects in the dissertation was performed in the following institutes and companies: Hungarian Academy of Sciences, Biological Research Centre; Roche Hungary Ltd.; Seqomics Ltd.. The genome projects were carried out with different next generation platforms, namely SOLiD (Life Technologies), Illumina MiSeq, and Roche GS Junior. The following bioinformatic software and applications were utilized: Genomics Workbench 4.9 (CLC Bio), Omixon Gapped SOLiD alignment 1.3.2, Mira V 4.0.2, NCBI Prokaryotic Genomes Automatic Annotation Pipeline, RAST (Rapid Annotation using Subsystem Technology).

In case of the K13 and *T. cellulositytica* TB100^T, the sequencing was performed with Illumina MiSeq. As a hybrid method, combination of data from SOLiD 4 and Roche GS Junior platform was used in the genome project of *T. halotolerans* JCM16012^T. The *T. fusca* TM51 full genome sequence is a result of a resequencing project. In this project the *T. fusca* YX genome sequence was used as a reference. The genome mining and the manual correction of the genome annotation was done with different online databases and software products: Pfam, Interpro, SignalP, TATfind, Expasy, and TMHMM.

2.4. Examination of *Thermobifida* endomannanases

During my examinations I worked with all four species belonging to the *Thermobifida* genus. The sources of the studied strains were different. *T. fusca* TM51 and *T. cellulositytica* TB100^T originated from the strain collection of our laboratory, but we had to purchase the *T. halotolerans* JCM16012^T and *T. alba*

CECT3323 strains (synonym: *T. alba* ULJB1) from the Japan Collection of Microorganisms and the Spanish Type Culture Collection, respectively.

2.4.1. Cloning of the endomannanase genes

Based on the genome projects data I could design 3 primer pairs for the *Thermobifida* endomannanase genes (Table 1.). Each forward primers harbour *NdeI* restriction site, and each reverse primers harbour *XhoI* restriction site. Beside the restriction sites there are 7-8 random bases at the 5' end of the primers to help the restriction endonuclease binding. I designed the primers for endomannanase genes without signal sequence coding nucleotide sequence. The position of signal sequence was determined with SignalP online application.

Table 1. Primers designed for the *man5A* genes (the underlined nucleotides are the restriction sites of the endonucleases)

Targeted gene	Primer sequences (5'-3')	
<i>T. fusca</i> endomannanase (<i>man5ATf</i>)	GGTGCCAT <u>CATATG</u> GCCACCGGGCTCC	forward primer
	GTGCCAT <u>TCTCGAGT</u> CAGCGAGCGGTG	reverse primer
<i>T. cellulositytica</i> endomannanase (<i>man5ATc</i>)	GGTGCCAT <u>CATATG</u> GCGACCGGGATCC	forward primer
	GTGCCAT <u>TCTCGAGT</u> CAGTCGACGGAGCAGGTC	reverse primer
<i>T. halotolerans</i> endomannanase (<i>man5ATh</i>)	GGTGCCAT <u>CATATG</u> GCCACCGGGCTTC	forward primer
	GTGCCAT <u>TCTCGAGT</u> CAGTCGGTGGTG	reverse primer

The amplification of endomannanase genes was performed with optimized PCR and *Pfu* polymerase. After the enzymatic digestion of the purified PCR products with *NdeI* and *XhoI* restriction endonucleases, the fragments were isolated from agarose gel with the appropriate kit (Wizard® SV Gel and PCR Clean-Up System, Promega GmbH, Mannheim). Plasmid (pET28) vector and T4 ligase were used for fragment ligation. After the ligation, *E. coli* Top10 competent cells were transformed with the vectors, and after culturing (12 hours, 37 °C, LB medium containing 50 µg/ml kanamycin) the plasmids were isolated. These

isolated plasmids were used to transform *E. coli* BL21 (DE3) cells and the protein overexpression was carried out with BL21 competent cells.

2.4.2. Protein expression and purification

Transformants were grown in LB medium containing 50 µg/ml kanamycin until optical density measured at 600 nm reached 0.6-1.0. Protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM final concentration), followed by overnight agitation at 20 °C. After the lysis of the biomass the supernatant was loaded onto a Hi Trap column (GE Healthcare) for immobilized metal ion affinity chromatography (IMAC) purification. Protein elution was performed with a 0-500 mM imidazole gradient in phosphate buffer, and protein concentration was determined by Bradford method (like in the subsequent cases). The successfulness of the protein expression and purification was checked by SDS-PAGE analysis.

2.4.3. Biochemical characterization of the endomannanases

The substrate specificity of the endomannanases was assayed using different polysaccharides (carboxymethyl-cellulose, crystalline cellulose, beechwood xylan, and galactomannan (locust bean gum)) and 4-nitrophenyl β-D-mannopyranoside (pNP-βM) as an artificial aryl-mannoside substrate.

The endomannanase activity was determined on polysaccharide substrate by measuring liberated reducing sugars according to dinitrosalicylic-acid method. The typical reaction mix of the different enzyme assays contained 450 µl phosphate buffer (50 mM, pH 7.5) in the presence of 2 mg/ml galactomannan and 50 µl enzyme solution (0.2-0.7 µg). The reaction took 10 minutes, and it was stopped by 1 ml Miller reagent.

The Michaelis-Menten kinetic parameters were determined in the presence of 0.4-4 mg/ml galactomannan substrate. The pH dependence on endomannanase activity was measured in the pH range of 4-10. The following buffers were used (pH ranges are indicated in brackets): 100 mM citrate-phosphate (pH 4.0-6.5),

100 mM sodium phosphate (pH 6.5-7.5), 200 mM Triethanolamine/HCl (pH 7.5-9.0) and 200 mM glycine/NaOH (pH 9.0-10.0). The effect of temperature on endomannanase activity was determined in 50 mM sodium phosphate buffer ($\text{Na}_2\text{HPO}_4\text{-NaH}_2\text{PO}_4$), pH 7.5 at different temperatures ranging from 40-90 °C. Thermal stability of the three endomannanases was assessed according to the following method: enzyme samples in 40 µg/ml of final concentration were incubated in phosphate buffer at 60 and 70 °C, from the incubated endomannanase solutions time course aliquots (10 µl) were withdrawn, cooled on ice for at least 30 min, and then assayed for endomannanase activity. The residual activity was calculated as a fraction of the initial activity and plotted against time.

3. Results

3.1 Building of compost originated strain collection

According to my first main objective, I have isolated bacteria strains on polysaccharide containing agarose plate and identified them based on the 16S rDNA sequences. As a result of my work, I have created/developed a strain collection consisting of 64 isolates belonging to 21 different species. Several members of the strain collection belong to *Geobacillus*, *Paenibacillus*, *Bacillus*, *Brevibacillus* genera and to the *Actinomycetales* order.

3.2 Taxonomical investigation of K13 strain

During the building of strain collection, a strain designated as K13 was isolated and after 16S rDNA analysis the highest similarity found was 93.03% in the EzTaxon database for *Paenibacillus montaniterrae* MXC2-2^T. Based on the phylogenetic tree of 16S rRNA genes, the strain K13 shows relationships/similarity with *Paenibacillus*, *Fontibacillus*, and *Cohnella* genera.

The strain K13 with extraordinary xylan-degrading capability is stick/rod shaped (2-3 μm long and 0.5-0.8 μm wide), Gram positive, aerobic, spore-forming, and mesophilic bacterium. The isolate can be cultured well in the neutral and alkaline range, the pH optimum is 9. According to the temperature optimum, the organism is mesophilic (40-45 $^{\circ}\text{C}$) and can be cultured in the presence of 2-4% salt. Based on the API20NE test, the strain is capable of the hydrolysis of esculin and gelatine and has β -galactosidase activity. The result of API50CH was completely negative (this test analyses the acid-producing capability), but presumably this means that the strain is not able of certain substrate metabolism under the given culturing conditions.

Regarding chemotaxonomic features, the dominance of MK-7 is typical for the strain and also the presence of meso-diaminopimelic acid in the cell wall. Regarding the polar lipid profile, which is determinative in related taxons, K13 contains the following polar lipids (in decreasing quantity): diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, and aminophospholipid. The components of membrane fatty acids are the followings: anteiso C15:0 (34.44 %), iso C16:0 (17.28 %), and C16:0 (9.96 %).

3.3 Results of genome analysis

The number of published genomes has greatly increased recently due to the spread of new-generation sequencing methods. In possession of the genome encoded information, mapping the metabolic activity of certain organism can be performed more effectively. The deeper genetical analysis based on the draft (providing partial data) genome sequences has its limitations, like a real physical map cannot be prepared, but they are suitable for searching for target genes or gene groups. During my work, my main aim was to explore the lignocellulose-degrading potential encoded in compost inhabiting bacteria by whole genome sequencing. The annotated draft genomes of the strain K13 and three *Thermobifida* species (*T. fusca*, *T. cellulositytica*, and *T. halotolerans*) gave the

opportunity to search for cellulase, xylanase, and mannanase genes that after heterologous expression were and will be targeted by different experiments.

The whole genome sequence of strain K13 was organised into 26 contigs. The genome is 4 251 364 bp and encodes 3 937 hypothetical proteins, from which I could successfully identify several enzyme genes belonging to the glycoside hydrolase (GH) family. During the annotation and manual data analysis, I found 33 GH enzyme genes in the genome belonging to 21 different families. The modular enzymes of the catalytic and carbohydrate binding domains are remarkably diverse. The strain K13 is able to produce various cellulose-binding domains, kadherin, immunoglobulin-like domain, and enzymes carrying fibronectin parts; additionally, the appearance of cellulases, xylanase, and mannanases containing (surface layer homology) SHL domain is rather interesting.

Summarising, K13 enzyme genes involved in the xylan-degradation have an outstandingly high number (15) and great variety. These genome data coincide with the high xylanase activity experienced during cultivation.

The type strain of the *Thermobifida* genus, *T. fusca*, which is the model organism of aerobic lignocellulose degradation, and other species of this genus (*T. cellulositytica*, *T. halotolerans*, *T. alba*) are proven to be extraordinary lignocellulose degraders. Three of the described species have been involved in genome projects (the *T. alba* genome sequencing is in progress). Due to the *T. fusca* TM51 genome project, the whole DNA data could be assembled into 88 contigs. Based on the results, the strain genome is 3 599 272 bp and contains 3 080 hypothetical proteins. The genome project of *T. fusca* YX strain was formerly published, so it was possible to compare the glycoside hydrolase genes of the two strains. The GH gene set of strain TM51 and *T. fusca* YX is identical, encoding 17 cellulase, xylanase, and mannanase glycoside hydrolases that belong to 12 different GH families. The two published *T. fusca* total DNA sequences, the genome of TM51 and YX strain, show 99.85% similarity.

Comparing the glycoside hydrolase genes of the two strains involved in lignocellulose degradation, I revealed 27 SNPs (single nucleotide polymorphism) and 13 amino acid exchanges that could cause the occasional differences between the enzymes' features.

Only one cutinase enzyme has been described from the *T. cellulosilytica* species, despite the fact that based on the genome project of *T. cellulosilytica* TB100^T, the gene set encoding the lignocellulose degrading enzymes is identical to the *T. fusca*. The published genome sequence provides the opportunity to study the *T. cellulosilytica* originated hydrolases. As a result of the genome project of TB100^T, the genome is of 4 327 869 base pairs and was assembled into 168 contigs and 19 scaffolds encoding 3 589 proteins.

The *T. halotolerans* is the only *Thermobifida* species that was not isolated from compost but from a salt mine, therefore it was interesting to compare the GH gene set of this species to other *Thermobifida* species. Based on the project, the genome of JCM16012^T is 4 123 689 bp and encodes 3 227 hypothetical proteins. The genes involved directly in the lignocellulose degradation show great similarity to the other two *Thermobifida* species. Enzyme genes of xylan and mannan hydrolases among the enzyme groups involved in lignocellulose degradation could be found in the *T. halotolerans* genome. Regarding cellulose enzymes, seven were found in the annotated genome of *T. fusca* and *T. cellulosilytica*, and five was identified in this strain. Three cellulase enzymes and one GH11 endoxylanase were described, but after the genome project, new enzymes' description can be expected.

3.4. Examination of endomannanases from three different *Thermobifida* species

The enzymes of the *Thermobifida* species have a significant industrial potential, due mainly to their high heat stability. Hitherto the publications focus on the *T.*

fusca cellulolytic enzyme system, although enzymes of the other *Thermobifida* species, including the hemicellulases, also have similar industrial potential.

During my work I carried out the comparison and description of homologous endomannanases originating from the four described *Thermobifida* species (*T. fusca*, *T. cellulositytica*, *T. halotolerans*, and *T. alba*). As the results of my examinations clearly indicated that the *T. alba* CECT3323 (it was ordered from the Spanish Type Culture Collection) was previously miss-classified and in the reality it is a *T. fusca* strain, it was excluded from the subsequent investigations. In the three investigated *Thermobifida* genome, the *man5A* endomannanase gene has an island-like location, near other glycoside hydrolases (1 endoglucanase and one β -mannosidase).

All three mannanase enzymes (originated from *T. fusca* - Man5ATf, from *T. cellulositytica* - Man5ATc, and from *T. halotolerans* - Man5ATh) have a modular structure. The endomannanase proteins contain N-terminal catalytic domain and C-terminal carbohydrate binding module (CBM2). There is a linker region between the domains, which are built up from 3-6 times repeated tetrapeptide motifs: 3xTEEP-Man5ATf, 5xPTDP-Man5ATc, and 6xDPGT-Man5ATh. The comparison revealed about 80% similarity among the amino acid sequences of the enzymes.

After the protein expression and purification, the SDS-PAGE indicated the molecular mass of about 50 kDa in all three cases. According to the substrate specificity assay, the proteins have endomannanase activity, and they are active only on galactomannan (locust bean gum) among the investigated substrates (carboxymethyl-cellulose, micro-crystalline cellulose, xylan, pNp-mannopiranozid).

Taking the pH and temperature dependence of enzymes into consideration, the investigated catalytic proteins have a thermal optimum at the 75-80 °C range, similarly to other *Thermobifida* extracellular enzymes. The pH optima of Man5A proteins are slightly alkaline (pH 7-8), that is also usual among the so

far described enzymes of the genus and among the compost inhabitant bacteria. The relative wide functional pH optimum range is also an important character of the investigated enzymes, particularly in case of Man5A.

The kinetic parameters of the proteins were determined according to the Michaelis-Menten model. Kinetic performances expressed in catalytic constant (k_{cat}) are similar for all three investigated mannanases (Man5ATf- $122\pm 11\text{ s}^{-1}$, Man5ATc- $89\pm 5\text{ s}^{-1}$, and Man5ATh - $78\pm 9\text{ s}^{-1}$), though results show that enzyme Man5ATf has the lowest affinity towards galactomannan substrate (highest K_M value).

Comparing the attributions of enzymes, the most marked differences were between heat stabilities. Despite the high sequence similarity, there are surprising differences between heat stabilities of the enzymes measured at 60 and 70 °C. During treatment at 60 °C, Man5ATf and Man5ATc enzymes maintain their activity for 3 hours, but the activity of Man5ATh decreases below 50% after 40 minutes. Only the Man5ATf could resist the heat treatment at 70 °C, while the other two endomannanases lose their active configuration after half an hour at this temperature. According to measurements, the most robust enzyme was Man5ATf, while the least heat resistant was Man5ATh.

3.5. New scientific results

Thesis I: I have isolated a previously unknown, new species, new genus candidate bacterium strain, and carried out its classic and molecular taxonomic description. The strain K13, isolated from compost, has an extremely high xylanase activity according to culturing studies. *Xylanobacillus xylanolyticus* sp. nov., gen. nov. K13 strain was deposited in two international bacterial culture collections (DSM 29793, NCAIM B.02605) as required for descriptions of new taxa.

Thesis II: Based on the *de-novo* genome project of *Xylanobacillus xylanolyticus* K13, I have identified 26 genes contributing to lignocellulose degrading

(cellulase, xylanase, and mannanase), belonging to 16 different glycoside hydrolase (GH) families. According to the numerous (15) and diverse (10 GH families) xylanase genes encoded in the genome, the strain K13 might become the model organism of aerobic, prokaryotic xylanase degradation.

Thesis III: I have made the genome project of *Thermobifida fusca* TM51 and the *de-novo* genome projects of *Thermobifida cellulositytica* TB100^T and *Thermobifida halotolerans* JCM16012^T, and based on them, the identification of all 17 glycoside hydrolase genes participating in lignocellulose degradation was carried out. Based on their comparison, it can be stated that GH gene sets of *T. fusca* TM51 and *T. cellulositytica* TB100^T are identical, while cellulase gene set of *T. halotolerans* JCM16012^T contains two more GH3 and one less GH6 and GH9 genes.

Thesis IV: I have proved by molecular taxonomical examinations that strain *T. alba* CECT3323 (earlier *Thermomonospora alba* ULJB1), specified as the source of hydrolases described from *Thermobifida alba*, in fact belongs to species *T. fusca*, thus there is no described hydrolase enzyme from *T. alba* at the present.

Thesis V: I was the first to clone *man5ATc*, originating from *Thermobifida cellulositytica*, and *man5ATh*, originating from *Thermobifida halotolerans*, genes, and after heterologous expression I have confirmed the endomannanase activity of their coded enzymes.

Thesis VI: I have determined biochemical parameters of Man5ATf, Man5ATc and Man5ATh enzymes, and I have found that the molecular weights of the proteins are 50 kDa, temperature optima are between 70-75 °C, and pH optima are in the slightly alkaline range (pH 7-8). By comparing the enzymes I have found that in spite of the amino acid sequence similarity of above 80%, the heat stability of Man5ATf endomannanase of strain *T. fusca* TM51 is outstanding,

but its enzyme affinity to galactomannan substrates is lower than of the other two endomannanases.

4. Conclusions, proposals

Members of the strain collection built with bacteria isolated from compost belong to specific genera and species characteristic of the lignocellulose degrading microbial community characteristic of composts. In a review, published in 2003, more than 150 different species present in compost were enumerated by literature data. The 64 isolates collected by me belong to 21 species, which is a high number regarding the conditions of selection (growth media with cellulose and hemicellulose contents, growth temperature of above 40 °C). Altogether, during the culturing work, I have found that increasing growth temperature decreases the number of species that can be isolated, and above 80 °C only species belonging to *Thermus composti* could be cultured.

Beside the building of strain collection, one of my main goals was the isolation, identification, and characterisation of new bacterium species. The 16S rDNS sequence of the species K13 shows the highest similarity to the 16S rRNS gene of the strain *Paenibacillus montaniterrae* MXC2-2^T, this value is only 93.03%, which is well below the 97% specified at the description of new species.

During the general tests applied for the determination of phenotypical characteristics and nutrient utilisation (API tests), under conditions by the manufacturer's instructions, no growth was observed for the K13 strain, so these results could not be evaluated.

Results of chemotaxonomical studies also revealed significant differences compared to the related species. K13 contains the diphosphatidylglycerol, phosphatidylethanolamine, and phosphatidylglycerol molecules characteristic to other groups of family *Paenibacillaceae* and to genus *Bacillus*, but as an individual trait, phosphatidylserine appears among polar lipids. In its fatty acid

composition, similar to genera *Fontibacillus*, *Cohnella*, and *Paenibacillus*, iso C_{15:0}, anteiso C_{15:0}, iso C_{16:0}, anteiso C_{17:0}, and C_{16:0} structured molecules are determinative. Results obtained during the characterisation of K13 strain confirm that it can be the type strain of a new species, new genus. Beside the formation of the new genus, the revision of the genus *Paenibacillus* may also be necessary to set up monophyletic groups. The guanine and cytosine (G+C) ratio of the DNA of the strain K13, or the presence of genes providing spore forming ability, was realized by genome projects, which shows that the total genome sequence analysis can contribute important data to taxonomic works as well.

During my work I have done total genome analysis of four species, primarily aiming at the mapping of glycoside hydrolase gene sets participating in lignocellulose degradation. In cases of *T. fusca* TM51, *T. cellulositytica* TB100^T, *T. halotolerans* JCM16012^T and strain K13, only (draft) genome sequences, providing partial data are currently available, which are only of limited sufficiency for deep genomics investigations, but for search of special target genes or mapping of metabolic potentials are perfectly suited.

Based on the genome sequences of the known three *Thermobifida* species, the GH gene sets of the members of the genus is quite similar. *T. fusca* and *T. cellulositytica* have the same genes encoding the enzymes of lignocellulose degrading. In case of *T. halotolerans*, in spite of the different habitat (the type strain has been isolated from a salt mine instead of compost, uniquely among the species of the genus), only small differences can be seen compared to the other two *Thermobifida* species. During the genome analysis of the three *Thermobifida* strains, 17 glycoside hydrolase genes, playing role in lignocellulose degrading could be identified.

By the culturing of K13 strain, the xylanase activity of the isolate was found outstanding, the genetic background of which was confirmed by the genome sequence analysis. Altogether 26 genes of lignocellulose degrading significance could be identified, among which the ratio of xylan degrading enzymes was

remarkably high (15), belonging to highly diverse GH families, so in the future, this strain could become the model organism of aerobic xylan decomposition.

Hereinafter, it would be worthwhile to prepare the physical map of the genomes as well, and with the more accurate genomic knowledge, to learn more details of the background of aerobic lignocellulose degradation, in particular of the controlling functions and the coordination of lignocellulolytic activities. Further research is needed for the exploration of actual roles of the diverse carbohydrate binding modules (CBM), and of the mechanisms promoting hydrolysis.

As the last phase of my work, I have characterised and compared the homologue endomannanase enzyme belonging to the GH5 family of three *Thermobifida* species. These were the Man5A_{Th} enzymes of *T. fusca*, *T. cellulositytica*, and *T. halotolerans*. My aim was also to study the endomannanase enzyme of *T. alba*, the fourth member of the genus, but the *T. alba* CECT3323 (or *T. alba* ULJB1) strain, ordered from the strain collection CECT, belonged to the species *T. fusca*, as it turned out. So it can be concluded that there is no described enzyme from the species *T. alba* at present, studies on XylA endoxylanase enzyme were in fact done on a *T. fusca* enzyme.

Genomic locations of the three studied enzyme genes are very similar, a cellulase (*cel5A*) and a mannosidase (*βman*) gene can also be identified near the *man5A* gene. The island-like arrangement of the hydrolase genes is a logical consequence of the fact that enzymes of various activities are needed for the degradation of lignocellulose with complex structure.

An above 80% similarity was found comparing the amino acid sequences of endomannanases, and also the modular structure characteristic of extracellular enzymes was proved by search in sequence databases. During the study of enzymes, analogous substrate specificity, and also similar pH and temperature optima were found. The Man5A enzymes were only active on galactomannan substrate from the studied polysaccharides and pNp-mannopyranoside. Regarding the pH optima, it was found that they are active in a relatively wide

and slightly alkaline range (pH 7-8). With this slightly alkaline pH optimum, thermobifida endomannanase enzymes represent a separate subgroup of the GH5 endomannanases. In respect of temperature dependence, it can be said for the three enzymes tested by me, that their temperature optimum is high (70-75 °C). According to the literature data, among eubacteria, only *Caldibacillus cellulovorans* and archaeon *Thermotoga neapolitana* produce endomannanase enzymes with significantly high temperature optimum (85 °C). According to the determined Michaelis-Menten kinetic parameters, the k_{cat} values of the three endomannanase enzymes are similar, though results show that enzyme Man5ATf has the lowest (1.65 ± 0.4 mg/ml) affinity, expressed in K_M value, to galactomannan substrate, while Man5ATc has a value of 0.84 ± 0.15 mg/ml. In the literature higher affinities are reported for endomannanases of *Aspergillus niger BK01* and *Bacillus sp. MG-33*. Majority of the publications report higher K_M values than received for Man5ATc.

Comparing the attributions of enzymes, the most marked differences were between heat stabilities. According to experiments carried out at 60 and 70 °C, the most heat resistant enzyme was Man5ATf, while the least robust was Man5ATh protein. The differences between heat stabilities mark the temperature optima of the studied *Thermobifida* strains as well, as the highest temperature optimum was reported for *T. fusca* TM51, while the lowest was found for *T. halotolerans* JCM16012^T.

It can be summarised that the three endomannanase enzymes studied are similar to one another in most biochemical aspects, however, despite the high degree of amino acid sequence similarity, striking differences were experienced in the thermal stability of proteins.

Previous studies show that changes in even a few amino acids might influence the important properties of some enzymes, thus, based on the results of the comparison, it would be possible to find those side chains that altered could increase heat stability of the given proteins.

Closely connected to my results presented in my dissertation, the experiments on the possibilities of practical use of the endomannanase of *T. fusca* TM51 strain have started last year.

6. Publications

Scientific paper:

Tóth Á, Baka E, Luzics Sz, Bata-Vidács I, Nagy I, Bálint B, Herczeg R, Olasz F, Wilk T, Nagy T, Kriszt B, Nagy I, Kukolya J. (2015) Plant polysaccharide degrading enzymes system of *Thermobifida cellulosilytica* TB100^T revealed by *de novo* genome project data. Acta Alimentaria DOI: 10.1556/066.2016.0014

Tóth Á, Barna T, Nagy I, Horváth B, Nagy I, Táncsics A, Kriszt B, Baka E, Fekete Cs, Kukolya J. (2013) Draft genome sequence of the lignocellulose decomposer *Thermobifida fusca* strain TM51. Genome Announc 1(4): e00482-13

Communication published in congress publications:

Tóth Á, Kriszt B, Baka E, Táncsics A, Nagy I, Horváth B, Stágel A, Fekete Cs, Kukolya J. (2013) Genome sequence of *Thermobifida halotolerans* type strain (YIM 90462T). Acta Microbiologica et Immunologica Hungarica 60 (Suppl.): p. 100. A Magyar Mikrobiológiai Társaság 2012. évi Nagygyűlése. Keszthely, Magyarország: 2012.10.24 -2012.10.26.

Luzics Sz, Tóth Á, Baka E, Varga S, la Riccia C, Bata-Vidács I, Szabó L, Schumann P, Tóth E, Kukolya J. (2015) *Xylanobacillus xylanolyticus* gen. nov., sp. nov., a moderately thermophilic species within the family *Paenibacillaceae*. Acta Microbiologica et Immunologica Hungarica 62 (Suppl.2): p. 179. 17th International Congress of the Hungarian Society for Microbiology. Budapest, Hungary: 2015.07.8. -2015.07.10.

Wilk T, Nagy T, Barta E, Olasz F, Bálint B, Herczeg R, Nagy I, Tóth Á, Luzics Sz, Kukolya J. (2015) De novo genome project of the polysaccharide degrader bacterium *Xylanobacillus xylanolyticus* gen. nov., sp. nov. *Acta Microbiologica et Immunologica Hungarica* 62 (Suppl.2): pp. 240-241. 17th International Congress of the Hungarian Society for Microbiology. Budapest, Hungary: 2015.07.8. -2015.07.10.

Tóth Á, Szabó E, Elek R, Hubert Á, Nagy I, Kriszt B, Táncsics A, Nagy I, Barna T, Kukolya J. (2015) Cloning, expression and biochemical characterization of endomannanases from three different *Thermobifida* species. *Acta Microbiologica et Immunologica Hungarica* 62 (Suppl.2): p. 231. 17th International Congress of the Hungarian Society for Microbiology. Budapest, Hungary: 2015.07.8. -2015.07.10.