Metagenomic cellulases from *Bacteria* linking IL-tolerance, halotolerance and thermostability

Dissertation
Zur Erlangung der Würde des Doktors der Naturwissenschaften
des Fachbereichs Biologie, der Fakultät für Mathematik, Informatik und Naturwissenschaften,
der Universität Hamburg

vorgelegt von
Nele Ilmberger
aus Hamburg

Hamburg 2011

Genehmigt vom Fachbereich Biologie
der Fakultät für Mathematik, Informatik und Naturwissenschaften
an der Universität Hamburg
auf Antrag von Prof. Dr. W. STREIT
Weiterer Gutachter der Dissertation:
Prof. Dr. B. SAAKE
Tag der Disputation: 26. August 2011

Hamburg, den 04. August 2011

Professor Dr. Axel Temming
Leiter des Fachbereichs Biologie

Table of Contents

Figures

Tables

Abbreviations

aa	amino acid(s)
Acc. No.	accession number
ad	up to
APS	ammonium persulphate
BAC	bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
bp	base pairs
BSA	bovine serum albumin
℃	degree Celsius
CBM	carbohydrate binding module
CMC	carboxymethylcellulose
Da	Dalton
DNA	deoxyribonucleic acid
DMF	dimethylformamide
DMSO	dimethyl-sulfoxide
DNSA	3,5-dinitrosalicylic acid
dNTP	deoxyribonucleotide triphosphate
DSMZ	German collection of Microorganisms and Cell Cultures (*"Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH"*)
EC number	enzyme commission number
EDTA	ethylene-diamine-tetraacetic-acid
et al.	et alii (Latin: and others)
Fig.	figure
g	gram(s)
g	radial centrifugal force
GC%	percentage of G and C in DNA sequences
GH	glycoside hydrolase
h	hour(s)
H-bond	hydrogen bond
IL	ionic liquid
IPTG	isopropyl thio-β-D-galactoside
IUBMB	International Union of Biochemistry and Molecular Biology
KAc	Potassium acetate
kb	kilobases
k_{cat}	catalytic constant
L	litre(s)

LB	Luria Bertani
µ	micro- (1×10^{-6})
m	milli- (1×10^{-3})
M	molar
mA	milliampere
Mbp	megabasepairs
min	minute(s)
MOPS	3-(N-morpholino) propanesulfonic acid
MSM	mineral salt medium
n	nano- (1×10^{-9})
NaCl	sodium chloride
Ni-TED	Nickel tris-carboxymethyl ethylene diamine
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gelelectrophoresis
PCR	polymerase chain reaction
pH	negative logarithm of the molar concentration of dissolved hydronium ions
rRNA	ribosomal RNA
RT	room temperature
SDS	Sodium dodecyl sulphate
sec	second
SSF	simultaneous saccharification and fermentation
subsp.	subspecies
TAE	Tris-acetate-EDTA
T_{ann}	annealing temperature
Taq	*Thermus aquaticus*
TE	Tris-EDTA
TEMED	N, N, N', N'-tetramethylethylenediamine
TIM	triosephosphate isomerase
T_m	melting temperature
$T_{opt.}$	optimum temperature
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
U	Unit
V	Volt
w/vol	weight per volume
vol.	volume
vol/vol	volume per volume
vol/w	volume per weight

X-Gal	5-Bromo-4-chloro-3-indolyl-β-D- galactopyranoside

Amino acids

Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

Nucleobases

A	Adenine
C	Cytosine
G	Guanine
T	Thymine

Ionic liquids

A) Cations

[AMIM]	1-allyl-3-methylimidazolium
[BMIM]	1-butyl-3-methylimidazolium
[BMMIM]	1-butyl-2,3-dimethylimidazolium

[BMPL]	1-butyl-1-methylpyrrolidinium
[EMIM]	1-ethyl-3-methylimidazolium
[MMIM]	1,3-dimethylimidazolium

B) Anions

[AC]	acetate
[ATF]	trifluoroacetate
[CL]	chloride
[DMP]	dimethylphosphate
[OTF]	trifluoromethanesulfonate

1 Introduction

1.1 Biotechnological relevance of cellulases

Nowadays, the need for energy sources alternative to fossil fuels is rising due to increasing oil prices, environmental problems, unhesitating exhaustion of fossil fuels and political instability in producing countries. Plant matter, i.e. starch and cellulose, can be used for the production of alternative energy. Therefore the respective substrate is hydrolysed to glucose which can be fermented to ethanol. These days, bioethanol usually is produced from sugarcane or cereals (Akoh et al. 2008) though this is in conflict with food supply. A possible alternative would be the generation of bioethanol from cellulosic material like wood or grass or from crop waste and forestry residues. The hydrolysis of cellulose can be performed either chemically (Bridgwater and Peacocke 2000) or enzymatically (Sánchez 2009). The chemical process proceeds via a combination of heat, pressure and acids and has as major disadvantages the high energy input and the formation of toxic waste products which have to be depolluted cost-intensively. The main drawback of the enzymatic process is that cellulose is not soluble in water and that the enzymes contradictory need an aquatic medium for a biotechnologically relevant conversion rate. Some ionic liquids (ILs) have been described to dissolve crystalline cellulose (Heinze et al. 2005, Swatloski et al. 2002). Therefore they could be used as an alternative reaction medium, if cellulases were active and stable in their presence. For this process two alternative approaches are possible. One possibility would be to directly apply the enzymes to the in IL dissolved cellulose. Therefore cellulases would have to be active in the presence of IL concentrations of nearly 100%. The other possibility is based on the fact that cellulose, which was dissolved in IL and that then becomes precipitated with water, has a reduced crystallinity and is therefore less recalcitrant to hydrolytic enzymes (Li et al. 2010). For this process, the enzymes would need to tolerate lower IL concentrations. But, disadvantageously, the reaction volume would be higher.

In addition to a possible role in the production of alternative fuels, cellulases are biotechnologically applied in the generation of the “used-look” effect of jeans (Bhat 2000), as additives in laundry agents (Ito 1997) and for the pulping of fruits that results in an improved extraction of juice (Bhat 2000). In paper industry cellulases are used for the de-inking of recycled fibres and for bio-mechanical pulping (Bhat 2000).

1.2 Ionic liquids (ILs)

Ionic liquids (ILs) are salts that are liquid at temperatures below 100 °C (Seddon 1997). They consist of an organic cation and an organic or inorganic anion (ions ILs are typically composed of are shown in Fig. 1.1). The formation of a stable salt crystal is sterically hampered because of the size and the symmetry of the ions. For industrial processes they have several advantages over organic solvents: they have no vapour pressure, are recyclable and they are liquid at temperatures over 300 to 400 °C, they are thermally, electrochemically and mechanically stable and they conduct electricity. Additionally ILs are able to dissolve a wide variety of substances (Seddon 1997). Some ILs, e.g. 1-butyl-3-methylimidazolium chloride ([BMIM]Cl) and 3-methyl-*N*-butylpyridinium chloride ([MBP]Cl), have been described to dissolve cellulose and/or wood and can be used to extract lignin from wood (Heinze et al. 2005, Kilpelainen et al. 2007, Lee et al. 2009, Swatloski et al. 2002). ILs can furthermore be designed individually for each reaction as ions can be paired arbitrarily and therefore features like water solubility, polarity, viscosity and melting point can be adjusted to the respective requirements. For these reasons, ILs are called “green solvents” or “designer solvents”.

For the mentioned reasons ILs are considered for a wide variety of reactions in chemical synthesis, i.e. hydration, oxidation, esterification and Friedel-Crafts-Reaction (Zhao 2002), in engineering processes (Zhao et al. 2005) and in enzymatic catalysis (van Rantwijk et al. 2003). Several lipases and esterases have been described to perform adequately in different ILs (Ganske and Bornscheuer 2005, Katsoura et al. 2006). In some cases, ILs even had beneficial effects on the stereoselectivity and/or enantioselectivity of the enzymatic reaction (Park and Kazlauskas 2001, Toshiyuki et al. 2006).

Additional to reactions which are performed in one phase, reaction systems with two or three phases are possible (Kragl et al. 2002). E.g. isoamyl acetate was synthesised at a microreactor scale in a two-phase system of n-heptane and 1-butyl-3-methylpyridinium dicyanamide with *Candida antarctica* lipase B (Pohar et al. 2009).

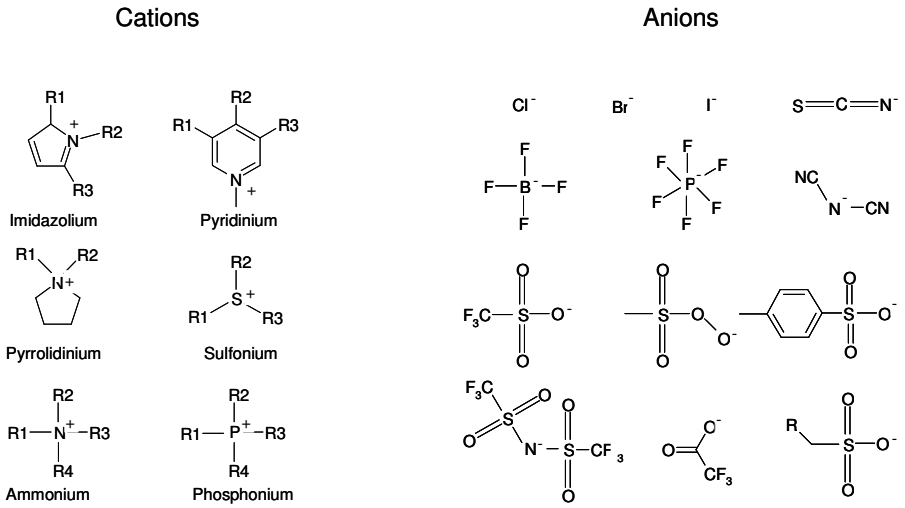

Figure 1.1: Commonly used ions for ILs

1.3 Cellulose

Cellulose is the world's most abundant biopolymer. It consists of β-1,4 linked D-glucose subunits, whereby every 2nd glucose is turned at an angle of 180° so that the dimer, cellobiose, is the smallest subunit of cellulose. As the glycosidic bonds in cellulose are bonds between C1 and C4, two different ends occur, the reducing and the non-reducing end of the cellulose strand (Fig. 1.2). Natural cellulose strands can reach a length of up to 15,000 glucose subunits. Via the formation of intra- and intermolecular hydrogen bonds, a very robust, crystalline structure occurs. Natural cellulose consists of crystalline and amorphous regions, whereby the degree of crystallinity can vary from 0 to 100% and depends on the frequency of hydrogen bonds. Cellulose is insoluble in water, most organic solvents and weak acids and bases. The half life time of crystalline cellulose at neutral pH and in the absence of microorganisms is about 100,000,000 years (Wilson 2008).

Naturally, the function of cellulose is to strengthen and to stabilise the plant and it therefore most often occurs in plant cell walls. The cellulose content of wood

depends on the plant and usually varies between 35 and 50% (dry weight). Other main components of wood are hemicellulose (~15%) and lignin (~30%). Also other plants, e.g. some grass species like switchgrass and *Miscanthus* sp. contain about 40% cellulose (Ververis et al. 2004). The cellulose content of cotton even reaches 90%. Next to plants, some bacteria and animals (*Tunicates*) can synthesise cellulose (Lynd et al. 2002).

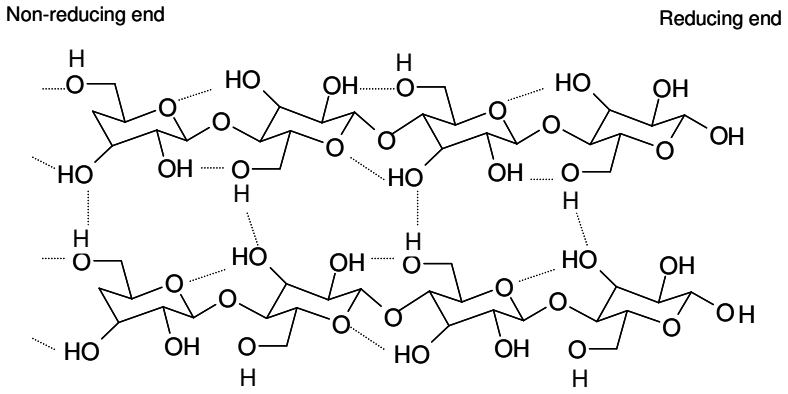

Figure 1.2: Structure of crystalline cellulose

Dotted lines: hydrogen bonds, solid lines: covalent bonds

1.4 Cellulases

Cellulases belong to the enzyme group of glycoside hydrolases (GH). These are enzymes which hydrolyse the glycosidic bonds between two carbohydrates or one carbohydrate and another variable group. Currently, more than 100 GH families are listed (http://www.cazy.org/Glycoside-Hydrolases.html). This classification depends on amino acid sequence similarities and therefore reflects the enzymes' structures but does not indicate substrate specifics as the IUBMB nomenclature does. This seems to be reasonable as many GHs are able to hydrolyse a wide variety of substrates (Liu et al. 2011, Warner et al. 2010, Yoon and Choi 2007). Some of the GH families can be grouped in 14 clans which are additionally distinguished by their fold. The hydrolysis of the glycosidic bonds proceeds via an acid hydrolysis mechanism using a proton donor and a nucleophile or base (glutamic acid or aspartic acid). The hydrolysis can either result in the inversion or the retention of the configuration of the anomeric C-atom (Sinnott 2002).

Three major types of enzyme activity are involved in cellulose hydrolysis. These are endoglucanase (EC 3.2.1.4), exoglucanase (EC 3.2.1.91) and cellobiase (EC 3.2.1.21) activity. Endoglucanases attack amorphous regions anywhere in the cellulose strand while exoglucanases liberate the glucose dimer, cellobiose, or larger oligomers from their ends, either from the reducing (Yaoi et al. 2004) or from the non-reducing end (Koivula et al. 1996). Cellobiases hydrolyse cellobiose to glucose.

Next to the catalytic domain, cellulases can contain further domains. Carbohydrate binding modules (CBM) mediate the attachment of the enzymes to the substrate and are necessary for the hydrolysis of crystalline cellulose (Bolam et al. 1998, Carvalho et al. 2004). Ig-like or Fn3-like domains have usually been found in enzymes belonging to GH family 9. The function of these domains is not completely elucidated, but the deletion of the Ig-like domain resulted in a complete loss of enzyme activity (Kataeva et al. 2002, Kataeva et al. 2004). Dockerin domains are found in cellulosomal enzymes. The majority of cellulases of anaerobic bacteria are organised in cellulosomes. These are large protein complexes consisting of a scaffold protein which is attached to the substrate via a CBM and furthermore linked to the bacterial cell wall (Doi et al. 2003). Different enzymes degrading the plant cell wall, e.g. cellulases, xylanases and laccases, are bound to the scaffold as the enzymes' dockerin domains specifically interact with the respective cohesin domains of the scaffold protein. With these protein complexes the organism is enabled to completely digest and/or to penetrate plant cell walls (Fontes and Gilbert 2010). Cellulosomes can result in a weight of up to 2.5 MDa; polycellulosomes that are composed of several cellulosomes can even reach 200 MDa (Doi et al. 2003). Genes coding for cellulosomal proteins are usually organised in gene clusters (Doi et al. 2003, Tamaru et al. 2000).

In contrast, aerobic bacteria and also anaerobes code for cellulases which are not organised in complexes but are functional as free proteins. These enzymes bind directly to the substrate. A study investigating chimeric enzymes has shown that enzymes either containing a CBM (free enzyme) or a dockerin domain (cellulosomal enzyme) show comparable activity (Vazana et al. 2010).

1.5 Activity of cellulases in the presence of ILs

Until now only few studies investigating the performance of cellulases in the presence of ILs have been published (Table 1.1).

Table 1.1: Traits of recently published cellulases that have been investigated in the presence of ILs

Cellulase	Specific activity (U/mg)	Relative activity (%) in (vol/vol %) IL	Residual activity (%) (incubation time) in (vol/vol %) IL	Organism	Reference
Cel5A	294	65 in 20% [BMIM]Cl	80 (5 h) 40% [BMIM]Cl	*Thermoanaero-bacter tengcongensis*	(Liang et al. 2011)
Celluclast® (Novozyme)	0.096	40 in 10% [MMIM][DMP]	40 (11 d) 10% [MMIM][DMP]	*Trichoderma reesei*	(Engel et al. 2010)
Cellulase (Fluka)	-*	50 in 20% [BMIM]Cl	87 (40 min) 10% [BMIM]Cl	*Aspergillus niger*	(Salvador et al. 2010)
T. viride	11	40 in 5% [EMIM]Ac	0 (15 h) 15% [EMIM]Ac	*Trichoderma viride*	(Datta et al. 2010)
Tma Cel5A	30	40 in 20% [EMIM]Ac	44 (15 h) 15% [EMIM]Ac	*Thermotoga maritima*	(Datta et al. 2010)
Pho EG	1.9	95 in 20% [EMIM]Ac	79 (15 h) 15% [EMIM]Ac	*Pyrococcus horikoshii*	(Datta et al. 2010)
Cellulase preparation	-	-	40 (1 d) 50% [BMIM]Cl, but 0 (1 d) 30% [BMIM]Cl	*Penicillium janthinellum*	(Adsul 2009)
CelA10	2.4	74 in 30% [BMPL][OTF]	0.8 (17 h) 60% [BMPL][OTF]	metagenome	(Pottkämper et al. 2009)
CelA24	23.3	2 in 30% [BMPL][OTF]	0.02 (17 h) 60% [BMIM][OTF]	metagenome	(Pottkämper et al. 2009)

*- no data available

Many of the investigated cellulases lost activity drastically in the presence of ILs (Pottkämper et al. 2009, Turner et al. 2003). A study investigating metagenomic cellulases from mesophilic habitats showed that most of the discovered enzymes lost their complete activity in the presence of 30% (vol/vol) IL, but one enzyme showed moderate activity in the presence of different ILs (Pottkämper

et al. 2009). Nevertheless, the thermostable cellulase from *Thermoanaerobacter tengcongensis* MB4 still showed 54.4% of its initial activity in 1.0 M (approx. 20% (vol/vol)) [BMIM]Cl (Liang et al. 2011). A thermophilic enzyme from *Pyrococcus horikoshii* was very active in the presence of 1-ethyl-3-methylimidazolium acetate [EMIM]Ac, it showed 95% relative activity in 20% (vol/vol) IL and 79% relative activity after 15 h in 15% (vol/vol) IL (Datta et al. 2010). Nevertheless, none of the enzymes showed high activity after an incubation of several days in the presence of ILs.

Altogether the industrial need for novel cellulases which tolerate specific conditions, i.e. the presence of high IL concentrations, and especially for those that are active over a long time period in the presence of ILs is rising.

1.6 Metagenomics

An appropriate tool for the discovery of novel enzymes is metagenomics. Metagenomics is the investigation of the "metagenome", the genomes of all microbes inhabiting a specific habitat. As this technology is culture-independent, the complete microbiota can be investigated instead of only the cultured fraction of 0.1 to 1% (Amann et al. 1995). Several methods are summarised under the term metagenomics, including 16S rRNA gene analysis, complete sequencing and the construction and investigation of metagenomic libraries (Fig. 1.3). 16S rRNA gene analysis is a powerful tool for the phylogenetic characterisation of habitats, e.g. the hydrolytic community in a switchgrass-adapted compost was described in comparison with the initial community. This study revealed drastic adaptations of the microbiota that was after 31 days dominated by a strain closely related to a genus (*Stackebrandtia*) whereof a cellulolytic representative is known (Allgaier et al. 2010). Other groups sequenced the complete metagenome of the chosen habitat, e.g. the termite gut microbiota was analysed (Warnecke et al. 2007). Within the recovered 71 Mbp GHs belonging to 45 families and more than 100 gene modules involved in cellulose degradation have been discovered. A commonly used tool for the discovery of novel enzymes and secondary metabolites like antibiotics is the construction and screening of metagenomic libraries. To construct metagenomic libraries, the total microbial DNA of the selected habitat

is isolated, ligated into suitable vectors (e.g. plasmid, cosmid, fosmid, BAC), and transferred into an appropriate host. The resulting metagenomic library can be used as fundament for ecological studies and can be screened for enzymes and other characteristics. This can be performed using function- or sequence-based screening approaches. Function-based screenings have the advantage that only functional enzymes are discovered and that, in contradiction to sequence-based screenings, also completely novel enzymes can be detected. Sequence-based screening approaches like PCR or hybridisation methods depend on the primer and probe design which relies on homologies of known sequences. Consequentially these methods can only be used to detect DNA fragments related to known sequences. The major advantage of sequence based-screenings is that the detection of enzymes does not depend on their correct expression, folding and secretion in the host bacterium. As in *Escherichia coli*, the most frequently used host bacterium, only about 40% of the bacterial enzymes can be detected by their function (Gabor et al. 2004a), alternative hosts for metagenomic libraries have been established. These are i.e. *Pseudomonas syringae* (Staskawicz et al. 1987), *Rhizobium leguminosarum* (Li et al. 2005) with a high number of secretion systems and *Thermus thermophilus* as a thermophilic host (Angelov et al. 2009).

In the past years a variety of novel enzymes have been detected with a metagenomic approach. This includes lipases (Elend et al. 2007), esterases (Elend et al. 2006), xylanases (Hu et al. 2008), amidases (Gabor et al. 2004b) and other characteristics like antibiotics (Gillespie et al. 2002) and operons for vitamin biosynthesis (Knietsch et al. 2003a). Also several cellulases with interesting properties have been detected. Pottkämper et al. (2009) described the identification of a novel cellulase which is moderately active in the presence of different ILs, and Voget et al. (2006) discovered a cellulase with high tolerance against NaCl.

The habitats that were mainly investigated and where most discovered enzymes derived from are highly diverse samples like soil samples and intestinal tracts (Daniel 2004, Elend et al. 2006, Feng et al. 2007, Schipper et al. 2009, Warnecke et al. 2007) but also extreme habitats like hot springs have been investigated (Park et al. 2007, Schoenfeld et al. 2008). Additional to the investigation of natural habitats, enrichment cultures can be advantageous as

microorganisms with the desired enzymes can be enriched and properties like temperature, pH, ionic conditions etc., can be adjusted. Furthermore DNA purification is facilitated as cultures contain fewer compounds that inhibit the isolation of high quality DNA, i.e. humic acids in soil samples (Schmeisser et al. 2007). Though microbial diversity is reduced (Daniel 2004, Knietsch et al. 2003b), several groups have been successful in the discovery of novel enzymes, e.g. dehydratases, from enrichment cultures (Knietsch et al. 2003a, Wang et al. 2008).

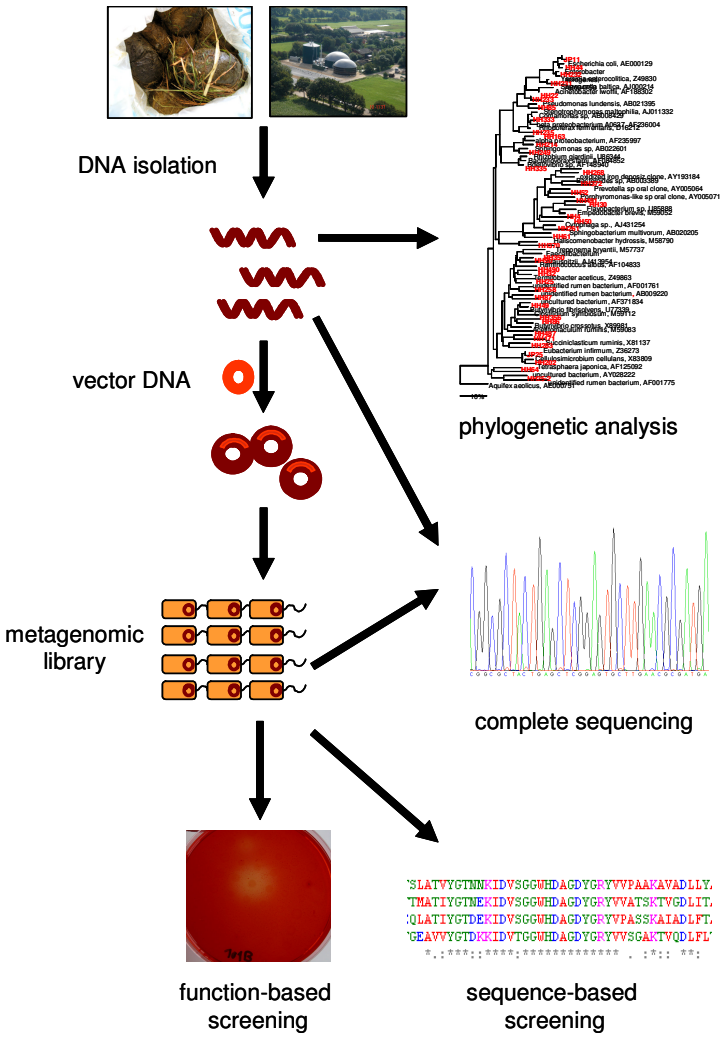

Figure 1.3: Metagenomic sample analysis

1.7 Intention of this work

The intention of this work was to identify and to characterise novel metagenomic cellulases which are active and stable in the presence of ILs.
Therefore different cellulolytic habitats including different enrichment cultures were evaluated regarding their biodiversity and their cellulolytic activity. From the most auspicious habitats metagenomic libraries were constructed and these were screened for clones with cellulolytic activity. The detected clones were investigated for their performance in the presence of ILs. The most promising enzymes were purified and characterised. As the enzymes that were most active in the presence of ILs were moderately thermophilic, two further thermostable enzymes were investigated with respect to their potential for biotechnological applications, these were CelA from *Thermotoga maritima* (Liebl et al. 1996) and Cel5K from a metagenomic library constructed with DNA from the Avachinsky Crater in Kamchatka.

2 Material & Methods

2.1 Bacterial strains, vectors, primers and constructs

In the following tables the bacterial strains (Table 2.1), vectors (Table 2.2), primers (Table 2.3) and constructs (Table 2.4) that were used in this study are listed.

Table 2.1: Bacterial strains used in this study

Strain	Characteristics	Reference/Source
E. coli DH5α	Cloning strain	Invitrogen (Karlsruhe, Germany)
E. coli EPI300	Host strain for pCC1FOS	Epicentre (Madion, WI, USA)
E. coli BL21 (DE3)	Strain for overexpression	Stratagene (La Jolla, CA, USA)
E. coli BL21 (DE3) pRIL	Strain for overexpression	Stratagene (La Jolla, CA, USA)
Ruminococcus albus (DSMZ 20455)	Cellulolytic strain	DSMZ (Braunschweig, Germany)
Clostridium cellulovorans (DSMZ 3052)	Cellulolytic strain	DSMZ (Braunschweig, Germany)

Table 2.2: Vectors used in this study

Vector	Characteristics	Reference
pDrive	Cloning vector	Qiagen (Hilden, Germany)
pBluescript SKII+	Cloning vector	Stratagene (La Jolla, CA, USA)
pCC1FOS	Fosmid for the construction of metagenomic libraries	Epicentre (Madion, WI, USA)
pET28a	Vector for the production of proteins with His-tag sequence	Stratagene (La Jolla, CA, USA)
pMAL-c2	Vector for the production of proteins with maltose binding protein (MBP)	New England Biolabs (Beverly, MA, USA)

Table 2.3: Primers used in this study

Primer	Sequence (5'-3')	Reference/Source
616V	AGAGTTTGATYMTGGCTCAG	(Juretschko et al. 1998)
1492R	CGGYTACCTTGTTACGAC	(Kane et al. 1993)
GM1F	CCAGCAGCCGCGGTAAT	(Muyzer et al. 1993)
M13 -20	GTAAAACGACGGCCAGT	Operon (Ebersberg, Germany)
M13 rev	CAGGAAACAGCTATGACC	Operon (Ebersberg, Germany)
T7 promotor	TAATACGACTCACTATAGGG	Novagene (Darmstadt, Germany)
T7 terminator	GCTAGTTATTGCTCAGCGG	Novagene (Darmstadt, Germany)
*Teredo*A	AATGGAATGCGGCACCA	this study
*Teredo*B	GAGCCAACGAACTGATGACA	this study
*Teredo*C	TTGGTTGCTGGCTGTGGAAGGGTA	this study
*Teredo*D	AGCCGGAGAACTTCAACCACATTG	this study
*Teredo*E	GACTGGTAGTAAGCATCCGCAGTT	this study
*Teredo*F	TGGTCAGTGTCCACAATG	this study
*Teredo*G	GTTGTTGGCGTTTCTGTTCC	this study
*Teredo*A-2	CTCAGCTCGCTATGCCAACG	this study
*Teredo*B-2	ACACCCATGAAGATACTGCG	this study
*Teredo*C-2	ATGAAGTCCATCATCCAGCG	this study
*Teredo*A-3	GCGCTGCGCCCATTGTTGC	this study
*Teredo*D-2	GTTTGCTGCGAAGCCTCGACC	this study
*Teredo*E-2	GTCTGACCTTTAGCGATTAG	this study
*celA84-Bam*HI-For*	GGATCCATGGATATATCTTATGGGA	this study
*celA84-Sal*I-Rev*	GTCGACACTTCAGGGTAATTGCCTG	this study
engB-For	TCATGGCTCCAGTATACAAGATAACCATT	this study
engB-Rev	TTAGAGGAACTGTTGTACTAGTGACTAA	this study
cel5D-For	AAGCGTACCACGCCGCACTCGCAAGAGTTA	this study
cel5D-Rev	AGGCAGGCTGTGAGAACATAGGAAGAGGAT	this study

*Recognition sites for restriction endonucleases are underlined

Table 2.4: Fosmids and constructs used in this study

Fosmid/construct	Characteristics	Reference
pFosCelA2	Metagenomic fosmid clone derived from a biogas plant sample	(Meske 2009)
pFosCelA3	Metagenomic fosmid clone derived from a biogas plant sample	(Meske 2009)
pFosCelA31	Metagenomic fosmid clone derived from an enrichment culture inoculated with an extract of *Teredo navalis*	This study
68D12	Metagenomic fosmid clone derived from elephant faeces	This study
82F10	Metagenomic fosmid clone derived from elephant faeces	This study
83D12	Metagenomic fosmid clone derived from elephant faeces	This study
pFosCelA84	Metagenomic fosmid clone derived from elephant faeces	This study
89C12	Metagenomic fosmid clone derived from elephant faeces	This study
89E1	Metagenomic fosmid clone derived from elephant faeces	This study
94A1	Metagenomic fosmid clone derived from elephant faeces	This study
101C1	Metagenomic fosmid clone derived from elephant faeces	This study
101D9	Metagenomic fosmid clone derived from elephant faeces	This study
172E9	Metagenomic fosmid clone derived from elephant faeces	This study
179F2	Metagenomic fosmid clone derived from elephant faeces	This study
pEngB	*engB* ligated into pDrive	This study
pCel5D	*cel5D* ligated into pDrive	This study
pET28a-*celA2*	*celA2* ligated into pET28	(Meske 2009)
pMAL-c2-*celA3*	*celA3* ligated into pMAL-c2	(Meske 2009)
pMAL-c2-*celA84*	*celA84* ligated into pMAL-c2	This study

2.2 Media and supplements

All media were autoclaved at 121°C for 20 min. Antibiotics and other heat sensitive supplements (Table 2.5) were sterile filtered and added to the media after these had cooled down to about 60°C.

2.2.1 Antibiotics and other supplements

The antibiotics and other supplements that were added to the media are listed in Table 2.5.

Table 2.5: Antibiotics and other supplements used in this study

Supplement	Final concentration in the medium	Concentration in the stock solution	Solvent
Ampicillin	100 mg/L	100 mg/mL	H_2O
Chloramphenicol	25 mg/L	25 mg/mL	Ethanol
Kanamycin	25 mg/L	25 mg/mL	H_2O
IPTG	100 mg/L	100 mg/mL	H_2O
X-Gal	50 mg/L	50 mg/mL	DMF

2.2.2 LB Medium

Agar (for agar plates)	15 g
Tryptone	10 g
NaCl	5 g
Yeast extract	5 g
H_2O *ad* 1000 mL	

2.2.3 Rich medium

Tryptone	10 g
NaCl	5 g
Yeast extract	5 g
Glucose	2 g
H_2O *ad* 1000 mL	

2.2.4 Mineral salt medium (MSM)

Buffer and mineral salt stock solution were autoclaved. After the solutions had cooled down, 100 µL of each stock solution were joined and 1 mL vitamin and 1 mL trace element stock solution were added. Then sterile water was added to a final volume of 1 L.

Buffer stock solution (10x)

Na_2HPO_4	70 g
KH_2PO_4	20 g
H_2O *ad* 1000 mL	

Mineral salt stock solution (10x)

$(NH_4)_2SO_4$	10 g
$MgCl_2 \cdot 6\ H_2O$	2 g
$Ca(NO_3)_2 \cdot 4\ H_2O$	1 g
H_2O *ad* 1000 mL	

Trace element stock solution (1000x)

EDTA	500 mg
$FeSO_4 \cdot 7\ H_2O$	300 mg
$CoCl_2 \cdot 6\ H_2O$	5 mg
$ZnSO_4 \cdot 7\ H_2O$	5 mg
$MnCl_2 \cdot 4\ H_2O$	3 mg
$NaMoO_4 \cdot 2\ H_2O$	3 mg
$NiCl_2 \cdot 6\ H_2O$	2 mg
H_3BO_3	2 mg
$CuCl_2 \cdot 2\ H_2O$	1 mg
H_2O *ad* 200 mL	

The solution was sterile filtered.

Vitamin stock solution (1000x)

Ca-Pantothenate	10 mg
Cyanocobalamine (B_{12})	10 mg
Nicotinic acid	10 mg
Pyridoxal-HCl (B_6)	10 mg
Riboflavin	10 mg
Thiamin-HCl (B_1)	10 mg
Biotin	1 mg
Folic acid	1 mg
p-Amino benzoic acid	1 mg
H_2O *ad* 100 mL	

The solution was sterile filtered.

2.2.5 Medium for *Clostridium cellulovorans*

$K_2HPO_4 \cdot 3\ H_2O$	1 g
NH_4Cl	1 g
KCl	0.5 g
$MgSO_4 \cdot 7\ H_2O$	0.5 g
Tryptone	0.5 g
Yeast extract	0.5 g
CMC	10 g
Na_2CO_3	1 g
$Na_2S \cdot 9\ H_2O$	0.15 g
H_2O *ad* 1000 mL	

After autoclaving, 1 mL trace element stock solution (see 2.2.3) was added to the medium.

2.2.6 Medium for *Ruminococcus albus*

Tryptone	5 g
Yeast extract	2 g
Glucose	3 g
Cellobiose	2 g
Mineral solution 1	40 mL
Mineral solution 2	40 mL
Na_2CO_3	4 g
Fatty acid mixture	1 mL
H_2O *ad* 1000 mL	

The pH was adjusted to 7.0 and the medium was then autoclaved.

Mineral solution 1:

K_2HPO_4	0.6 g
H_2O *ad* 100 mL	

Mineral solution 2:

KH_2PO_4	0.6 g
$(NH_4)_2SO_4$	2 g
NaCl	1.2 g
$MgSO_4 \cdot 7\ H_2O$	0.25 g
$CaCl_2 \cdot 7\ H_2O$	0.16 g
H_2O *ad* 100 mL	

Fatty acid mixture:

Isobutyric acid	1 mL
Isovaleric acid	1 mL
2-Methylbutyric acid	1 mL
H_2O *ad* 100 mL	

2.3 Samples, metagenomic libraries, cellulolytic clones and cellulases

2.3.1 Environmental samples

The following environmental samples were used directly and/or for enrichment cultures.

2.3.1.1 Elephant faeces

The elephant faeces sample used in this study derived from Kandy, an Asian elephant born 2003, living at *Hagenbecks Tierpark* (Hamburg, Germany). The sample was taken when the animal was 5 years old.

2.3.1.2 Koala faeces

The koala faeces sample derived from various animals from the zoo in Duisburg (Germany).

2.3.1.3 *Teredo navalis*

Teredo navalis (adult animals) derived from the North Sea.

2.3.1.4 Biogas plant

The biogas plant sample derived from a private plant in Linnau, Germany. The plant was fed with corn silage and run at a temperature of approximately 55°C.

2.3.2 Metagenomic libraries

The metagenomic library constructed with DNA of the elephant faeces was constructed by J. Juergensen and U. Rabausch (AG Streit, *Universität Hamburg*) and the library constructed with DNA isolated from the biogas plant by D. Meske (AG Schmitz-Streit, *Christian Albrechts Universität zu Kiel*). These libraries comprised 20,000 clones for the library derived from elephant faeces and 9,600 clones for the library derived from the biogas plant. The insert frequency was 100% for both libraries and the average size of the inserts was 37.5 kb for the library constructed with DNA from elephant faeces (Juergensen and Rabausch, pers. comm.) and 23.5 kb for the library constructed with DNA from the biogas plant (Meske 2009).

2.3.3 Cellulolytic clones

The biogas plant library was screened for cellulolytic clones by D. Meske (AG Schmitz-Streit, *Christian Albrechts Universität zu Kiel*). Positive clones were evaluated in this work for their performance in the presence of ILs. Thereby two clones, pFosCelA2 and pFosCelA3, were chosen for further investigation. Gene identification and cloning was performed by D. Meske (AG Schmitz-Streit, *Christian Albrechts Universität zu Kiel*). The gene *celA2* (Acc. No. JF826524) encoded a cellulase of 604 aa, including an glycoside hydrolase (GH) domain family 9 and an Ig-like domain. CelA2 showed 41% identity to EngO from *Clostridium cellulovorans* (55% similarity). The GH encoded on pFosCelA3 (634 aa, Acc. No. JF826525) included two C-terminal CBMs (family 17_28) and was

closest related to a GH5 cellulase from *C. cellulolyticum* (60% identity/75% similarity). CelA3 additionally contained a signal peptide (probability 0,999; SignalP-HMM). The pH optimum was determined as 6.5 for both cellulases (Meske 2009). The fosmid clones and the clones for overexpression (pET28a-*celA2* and pMAL-c2-*celA3* in *E. coli* BL21 (DE3) pRIL) were received from D. Meske for further investigation in this study.

2.3.4 Cellulases

The purified proteins CelA (Acc. No. CAA93273.1) and Cel5K (Acc. No. JF802029) were obtained from M. Mientus (AG Liebl, *Technische Universität München*). CelA was a GH family 12 endoglucanase from *Thermotoga maritima* and was previously described (Liebl et al. 1996). Cel5K derived from a metagenomic library constructed with DNA isolated from a soil sample derived from the Avachinsky Crater in Kamchatka. It was a GH5 cellulase with highest identity (61%) to *Thermus caldophilus*. Neither CelA nor Cel5K contained any additional domain. The temperature optima were 83°C for CelA (Liebl et al. 1996) and 96°C for Cel5K (Mientus, pers. comm.).

2.4 Cell culture, growth conditions and strain collection

2.4.1 Enrichment cultures

Enrichment cultures were grown in mineral salt medium (2.2.4) which was supplied with different cellulosic substrates, i.e. carboxymethylcellulose (CMC) or filter paper, as sole carbon source. The cultures were inoculated with faeces (2.3.1.1, 2.3.1.2), the biogas plant sample (2.3.1.4) or with *T. navalis* (2.3.1.3), which was minced in an ULTRA TURRAX® T18 basic homogenizer (IKA WORKS Inc., Wilmington, NC, USA) before inoculation. All enrichment cultures were gently agitated. Cultures inoculated with elephant faeces, koala faeces and the biogas plant sample were incubated at 37°C under microaerobic conditions, while the cultures inoculated with *T. navalis* were incubated aerobically at 22°C. The cultures were re-inoculated in new medium every two weeks.

2.4.2 Cultivation of bacterial strains

E. coli strains were grown, if not mentioned elsewise, at 37°C on LB medium (2.2.2) with appropriate antibiotics if necessary. Liquid cultures were shaken at 200 rpm.

Clostridium cellulovorans and *Ruminococcus albus* were grown under anaerobic conditions at 28°C in the respective medium (2.2.5, 2.2.6).

2.4.3 Determination of cell density

For the determination of the cell density of liquid cultures, the optical density (OD) was measured at a wavelength of 600 nm (OD_{600}) with a SmartSpec™ Plus spectrophotometer (BIO-RAD, Hercules, CA, USA). Cell numbers can be calculated based on the fact that an OD_{600} of 0.1 corresponds to a cell number of $1x10^8$ *E. coli* cells/mL.

2.4.4 Strain maintenance

For the long-term storage of bacterial strains, glycerol was added to overnight cultures to a final volume of 33% (vol/vol). Glycerol stocks were stored at -70°C.

2.5 DNA purification

2.5.1 Isolation of genomic DNA

For the isolation of (meta-) genomic DNA, cells were harvested by centrifugation at 13,000 *g* and 4°C for 5 min. The pellet was suspended in 1 mL washing solution and incubated for 1 h on ice. The suspension was centrifuged at 9,000 *g* and 4°C for 5 min and the cell pellet was resuspended in 250 µL TE-sucrose buffer. After the addition of 250 µL cell lysis buffer and subsequent mixing, the sample was incubated at 37°C for 1 h. Then 150 µL Proteinase K solution were added and the reaction mixture was incubated at 37°C for 1 h. After the addition of 250 µL phenol/chloroform (1:1, vol/vol) the sample was mixed and centrifuged at 13,000 *g* (15 min, 4°C) to induce phase separation. The supernatant was then transferred into a new sterile tube. The addition of phenol/chloroform with subsequent mixing and centrifugation was repeated until no protein layer occurred as an interphase. Afterwards, 250 µL chloroform were added and the sample was mixed. This was followed by another centrifugation

step at 13,000 *g* and 4°C for 5 min. The supernatant was transferred into a new sterile tube. To precipitate DNA, 2.5 vol. ethanol (99%) and 0.1 vol. 3 M sodium acetate (pH 5.5) were added. The tube was then inverted and incubated at -20°C for at least 1 h. Then the DNA was sedimented at 13,000 *g* and 4°C for 20 min and the supernatant was decanted carefully. The pellet was washed twice with 300 µL ethanol (70%, vol/vol) followed by a centrifugation of 2 min at 4°C and 13,000 *g*. After discarding the supernatant, the pellet was air-dried. The DNA was then dissolved in an appropriate volume of sterile H_2O. The integrity and quantity of the isolated DNA were analysed with spectrophotometry (2.6) and agarose gel electrophoresis (2.7).

Washing solution

NaCl	800 mM
EDTA	100 mM

The pH was adjusted to 8.0 and the solution was sterile filtered.

TE-sucrose buffer

Tris-HCl	10 mM
EDTA	1 mM
Sucrose	20% (w/vol)

The pH was adjusted to 8.0 and the solution was sterile filtered.

Cell lysis buffer

Tris-HCl	10 mM
EDTA	1 mM
Lysozyme	10 mg/mL
RNase A	1 mg/mL

The pH was adjusted to 8.0 and the solution was sterile filtered. RNase A was added after it was incubated for 15 min at 95°C.

Proteinase K solution

Proteinase K	1 mg/mL
Sarcosyl	5% (w/vol)

The sterile filtered solution was stored at -20°C.

2.5.2 Plasmid isolation with the QIAGEN® Plasmid Midi Kit

For the preparation of large amounts of plasmid DNA, the QIAGEN® Plasmid Midi Kit (QIAGEN, Hilden, Germany) was used. The buffers were supplied within the kit and used according to the instruction manual.

The cells of an overnight culture of 100 mL were harvested by centrifugation at 4,500 *g* and 4°C for 10 min. The pellet was suspended in 5 mL buffer P1 and the cell suspension was portioned to aliquots of 500 µL. To each aliquot 500 µL buffer P2 were added and the tubes were inverted several times with a subsequent incubation step of 5 min at RT. Then 500 µL buffer P3 were added to each tube and the tubes were inverted several times. After incubating the samples 15 min on ice, they were centrifuged at 13,000 *g* and 4°C for 10 min. The supernatant was transferred onto a QIAGEN®-tip 100 that was equilibrated with 4 mL buffer QBT. Afterwards, the plasmid DNA was washed 4 times with 5 mL buffer QC. Then the DNA was eluted from the column with 10 times 1 mL buffer QF. To each 1 mL portion 700 µL 2-propanol were added to precipitate the DNA. Then, after a centrifugation step at 13,000 *g* and 4°C for 30 min, the supernatant was discarded and the DNA pellet washed twice with 300 µL ethanol (70%, vol/vol) and centrifuged for 2 min. The pellet was air-dried before being dissolved in an appropriate volume of H_2O. The DNA quantity and integrity was determined by spectrophotometry and agarose gel electrophoresis (2.6, 2.7).

2.5.3 Plasmid isolation with the QIAprep® Spin Miniprep Kit

Plasmid isolations from small culture volumes were performed with the QIAprep® Spin Miniprep Kit (Qiagen, Hilden, Germany). All buffers were provided within the kit and were used as described in the respective manual. All centrifugation steps were performed at 13,000 *g*.

Cells from an overnight culture (4 mL) were harvested by a centrifugation step of 30 sec. The pellet was suspended in 250 µL buffer P1 before 250 µL buffer P2 were added. Then the tube was inverted 4 to 6 times before 350 µL buffer N3 were added. The tube was then inverted 4 to 6 times again and subsequently centrifuged for 10 min. Afterwards the supernatant was transferred onto the spin column and centrifuged for 1 min. The flow-through was discarded and the column was first washed with 500 µL buffer PB and then

with 750 µL buffer PE, each followed by a centrifugation step of 1 min. The flow-through was discarded. To remove residual buffer, the column was centrifuged again for 1 min. Then the DNA was eluted in 50 µL sterile H_2O via an incubation step at RT for 1 min and subsequent centrifugation for 1 min. DNA quantity and integrity were determined by spectrophotometry (2.6) and agarose gel electrophoresis (2.7).

2.5.4 "Quick and Dirty Prep"

The "Quick and Dirty Prep" was used to isolate plasmid and fosmid DNA. All centrifugation steps were performed at 13,000 *g* and RT if not mentioned elsewise.

An overnight culture of 4 mL was centrifuged for 30 sec and the sedimented cells were suspended in 100 µL buffer P1 before 200 µL buffer P2 were added. The tube was inverted several times and incubated at RT for 1 to 5 min until the clearance of the sample. Then 200 µL chloroform were added and the sample was mixed well. After the addition of 150 µL buffer P3, the tube was mixed and centrifuged for 5 min. The upper phase was transferred into a new tube and the same volume 2-propanol was added. After inverting, the sample was incubated at -20°C for 30 min. Then the sample was centrifuged at 4°C for 20 min. The supernatant was removed and the pellet washed twice with 300 µL 70% (vol/vol) ethanol, followed by a centrifugation step at 4°C for 2 min. Finally, the pellet was air-dried and suspended in 50 to 100 µL sterile H_2O. DNA quantity and integrity were determined with spectrophotometry (2.6) and agarose gel electrophoresis (2.7).

P1 buffer

EDTA	10 mM
Tris-HCl	50 mM
RNase A	1 mg/mL

The pH was adjusted to 8.0 and the solution was sterile filtered and stored at 4°C. RNase A was added to the autoclaved buffer after it was incubated for 15 min at 95°C.

P2 buffer

NaOH	200 mM
SDS	1% (w/vol)

The solution was sterile filtered.

P3 buffer

KAc	3 M

The pH was adjusted to 5.5 with acetic acid and the solution was sterile filtered.

2.5.5 Gel extraction of DNA with the QIAquick® Gel Extraction Kit I

To purify single DNA bands from agarose gels (2.6), the QIAquick® Gel Extraction Kit I (Qiagen, Hilden, Germany) was used. All buffers were supplied within the kit and used as described in the instruction manual. Centrifugation steps were carried out at 13,000 *g* and RT for 1 min.

The target DNA fragment was excised from the agarose gel (2.6) with a scalpel and transferred into a sterile tube. To 1 gel volume, 3 vol. buffer QG were added. The sample was incubated at 50°C until the complete dissolution of the gel slice was achieved. One gel volume 2-propanol was added and the sample loaded onto a QIAquick® column. After centrifugation, the flow-through was discarded and 500 µL buffer QG were transferred onto the column, followed by another centrifugation step. Buffer PE (750 µL) was added onto the column and the sample was centrifuged. Then the column was centrifuged again to remove residual buffer. The column was transferred into a sterile tube and the DNA was eluted in 30 to 50 µL sterile H_2O by incubation for 1 min at RT and subsequent centrifugation. The DNA concentration and integrity were determined by spectrophotometry (2.6) and agarose gel electrophoresis (2.7).

2.5.6 Purification of DNA fragments with the PCR clean-up Gel extraction Kit

DNA fragments were purified using the PCR clean-up Gel extraction Kit from Macherey-Nagel (Düren, Germany) according to the manufacturer's protocol. The buffers were supplied within the kit. All centrifugation steps were carried out for 1 min at RT and 13,000 *g*.

One sample volume and two volumes buffer NT were mixed and transferred onto the column. After centrifugation, the column was washed with the addition of 700 µL buffer NT3 and subsequent centrifugation. The column was dried with a further centrifugation step. Purified DNA fragments were eluted in 30 to 50 µL sterile H_2O, analysed via agarose gel electrophoresis (2.7) and stored at 4°C until further use.

2.6 Spectrophotometrical determination of DNA concentration and purity

To determine the quantity and purity of dissolved DNA, a SmartSpec™ Plus spectrophotometer (Bio-Rad, Hercules, CA, USA) was used. The extinction was measured at 260 nm in UV cuvettes (Brand, Wertheim, Germany). The DNA content can be calculated as an extinction of 1.0 corresponds to a concentration of 50 µg/mL. The purity of DNA is indicated by the ratio of the extinctions at 260 nm and 280 nm. Pure DNA has a $ratio_{260/280}$ of 1.8 to 2.0.

2.7 Agarose gel electrophoresis

The size, quantity and integrity of DNA were analysed with agarose gel electrophoresis. Agarose gels (0.8% (w/vol) in TAE buffer) were applied in an electrophoresis gel chamber (Hoefer™ HE-33 mini horizontal submarine unit, Amersham Biosciences, Piscataway, NJ, USA) filled with TAE buffer. Samples were supplied with 1/10 vol. loading dye before loading onto the gel. DNA fragments were separated at 120 V for 25 min with a power supply EPS 301 (Amersham Biosciences, Piscataway, NJ, USA). Agarose gels were stained in an ethidium bromide solution (10 µg/mL) for 15 min and washed briefly in a water bath to remove surplus ethidium bromide. Visualization and documentation were carried out in a Universal Hood II (BIO-RAD, Milan, Italy) supported by Quantity I 1-D-Analysis software (BIO-RAD, Philadelphia, PA, USA).

The determination of the size of DNA fragments was performed by the comparison with the standard GeneRuler™ 1 kb DNA Ladder (MBI Fermentas, St. Leon-Rot, Germany) that was also applied on the agarose gel.

TAE buffer (50x)

Tris	2 M
EDTA	100 mM

The pH was adjusted to 8.2 with acetic acid.

Loading dye

Glycerol	60 mL
EDTA	50 mM
Bromphenol blue	0.5 g
Xylencyanol	0.5 g
H_2O *ad* 200 mL	

2.8 Polymerase chain reaction (PCR)

Specific DNA fragments were amplified with PCR. PCR products were analysed using agarose gel electrophoresis (2.7).

2.8.1 PCR primers

Primer annealing was performed at the respective annealing temperature (T_{ann}) of the applied primers (2.1). The melting temperature T_m was calculated with the following equation (Chester and Marshak 1993).

$$T_m = 69.3\,°C + 0.41\,°C \times [GC\%] - (650 / \text{bp-length}_{Primer})$$

The annealing temperature T_{ann} was calculated as follows:

$$T_{ann} = T_m - 5\,°C.$$

For the PCR reaction the lower T_{ann} of both primers was used.

2.8.2 PCR conditions

PCRs were performed either in a Mastercycler personal (Eppendorf, Hamburg, Germany) or a Mastercycler gradient (Eppendorf, Hamburg, Germany). Table 2.6 illustrates the applied PCR reaction conditions.

Table 2.6: PCR reaction conditions

Initial denaturation		95°C	2 min
	Denaturation	95°C	30 sec
35x	Annealing	T_{ann} (2.8.1)	30 sec
	Elongation	72°C	1 min/kb
Final elongation		72°C	5 min

2.8.3 PCR volumes

The standard PCR reaction contained the following ingredients.

Template	1 µL
Taq polymerase buffer (10x)	2.5 µL
dNTPs (2 mM)	2 µL
Forward Primer (10 µM)	1 µL
Reverse Primer (10 µM)	1 µL
Taq polymerase	0.5 µL
H_2O *ad* 25 µL	

Taq polymerase buffer (10x)

Tris	100 mM
$MgCl_2$	25 mM
KCl	500 mM

The pH was adjusted to 8.3 and the solution was sterile filtered.

2.8.4 Direct colony PCR

Direct colony PCR was applied to verify putative positive clones after ligation (2.9.3) and transformation (2.10). Therefore specific vector primers (2.1) were used. The respective colonies were suspended individually in 20 µL water; 1 µL of this suspension was added to the PCR reaction as template. The standard reaction conditions (2.8.2) and volumes (2.8.3) were used.

2.9 Enzymatic modification of DNA

2.9.1 Site specific digestion of DNA

For the site specific digestion of DNA restriction endonucleases were applied. The following reaction mixtures were prepared A) for analytical and B) for preparative digestions.

A) Analytical digestion

DNA solution	1- 2 µL
Reaction buffer (10x)	1 µL
Restriction enzyme	0.5 µL
H_2O *ad* 10 µL	

B) Preparative digestion

DNA solution	5-10 µL
Reaction buffer (10x)	5 µL
Restriction enzyme	2 µL
H_2O *ad* 50 µL	

Analytical digestions were incubated for 1.5 h and preparative digestions for 3 h to overnight at the respective optimal temperature that was provided by the enzymes' deliverer, Fermentas (St. Leon-Rot, Germany). The digestions were analysed by agarose gel electrophoresis (2.7). Vector preparations were dephosphorylated (2.9.2) after digestion.

2.9.2 Dephosphorylation of complementary ends

To avoid religation of linearised vectors (2.9.1), these were dephosphorylated with Antarctic Phosphatase (New England Biolabs, Beverly, MA, USA). The reaction was set up as follows.

Preparative digestion	50 µL
Reaction buffer (10x)	6 µL
Antarctic Phosphatase	1 µL
H_2O	3 µL

In dependence of the manner of digestion of the respective enzyme (blunt ends or overlap), the reaction was incubated 15 min (5'-overlap) or 60 min (blunt ends and 3'-overlap) at 37°C. Afterwards, the enzyme was heat inactivated at 65°C for 15 min and the vector was purified (2.5.6).

2.9.3 Ligation of DNA

2.9.3.1 Ligation into pDrive

To ligate PCR products that were amplified with *Taq* polymerase (2.8), the QIAGEN® PCR Cloning Kit (QIAGEN, Hilden, Germany) was used. Purified

PCR products (2.5.6) were ligated into the pDrive cloning vector (2.1) within the following reaction.

PCR product	0.5 - 2 µL
pDrive vector	0.5 µL
Ligation master mix	2.5 µL
H_2O *ad* 5 µL	

The reaction was incubated at 16°C for 2 h and the ligation was then transformed (2.10) into *E. coli* (2.1).

2.9.3.2 Ligation into vectors with ends prepared with restriction enzymes

The ligation of ends generated by restriction enzymes (2.9.1) was carried out in the following reaction mixture.

Insert	0.2 µg
Vector	molar ratio insert:vector 2:1 to 10:1
Ligation buffer (10x)	2 µL
T4 DNA ligase (Fermentas)	0.5 µL
H_2O *ad* 20 µL	

The reaction was incubated for 2 to 16 h at 22°C and the ligation was then transformed (2.10) into *E. coli* (2.1).

2.10 Heat shock transformation

Plasmids and ligations (2.9.3) were transferred into *E. coli* via heat shock transformation.

2.10.1 Production of chemically competent cells

For heat shock transformations, chemically competent cells were produced. Therefore 250 mL LB (2.2.2) were inoculated 1:100 with an overnight culture of an appropriate *E. coli* strain (2.1). The culture was incubated at 37°C until an OD_{600} of 0.5 was reached. Then the culture was cooled 5 min on ice and centrifuged for 5 min at 4°C and 4,500 *g*. The supernatant was discarded and the cell pellet carefully suspended in 75 mL ice cold TFB1. After an incubation step for 90 min on ice, the cells were harvested again for 5 min at 4°C and

4,500 *g*. The supernatant was discarded and the pellet carefully suspended in 10 mL ice cold TFB2. The cells were stored as aliquots of 100 µL at -70°C until further use.

TFB1

RbCl	0.91 g
$MnCl_2 \cdot 4\ H_2O$	0.74 g
KAc	0.22 g
$CaCl_2 \cdot 2\ H_2O$	0.11 g
Glycerol (86%)	13 mL
H_2O *ad* 100 mL	

The pH of 5.8 was adjusted with acetic acid and the solution was sterile filtered.

TBF2

MOPS	0.042 g
RbCl	0.0242 g
$CaCl_2 \cdot 2\ H_2O$	0.221 g
Glycerol (86%)	3.5 mL
H_2O *ad* 20 mL	

The pH of 6.8 was adjusted and the solution was sterile filtered.

2.10.2 Transformation

For heat shock transformations, 100 µL competent *E. coli* cells (2.10.1) were thawed on ice. After the addition of 5 µL ligation (2.9.3) or plasmid the cells were incubated for 30 min on ice. The heat shock was performed at 42°C for 90 sec, before the cells were incubated on ice for 5 min. Then 800 µL LB medium (2.2.2) were added and the sample was incubated 45 min at 37°C. The cells were spread on agar plates containing the appropriate antibiotic and 100 µM IPTG and 40 µg/mL X-Gal (2.2.1), if necessary. The plates were incubated overnight at 37°C. Individual clones were analysed via direct colony PCR (2.8.4) and plasmid purification (2.5.4) with subsequent analytical digestion (2.9.1).

2.11 Sequence Analysis

2.11.1 Sanger sequencing

Sequencing was realised by the group of Prof. Schreiber at the *Institut für klinische Molekularbiologie* at the *Universitätsklinikum Schleswig-Holstein* in Kiel, Germany with an ABI 3730XL DNA Analyser (Applied Biosystems, Carlsbad, CA, USA) based on the Sanger technique (Sanger et al. 1977).
Plasmids were purified (2.5.3) and adjusted to a concentration of 100 ng/µL. Appropriate primers (2.1) were diluted to 4.8 µM in H_2O. Template (3 µL) and primer (1 µL) were joined and the sample then sent to Kiel.

2.11.2 Pyrosequencing of fosmids

To obtain the complete sequence of fosmids, they were sequenced via pyrosequencing. Therefore the respective fosmids were isolated (2.5.4) and analysed for their integrity, purity and quantity (2.6, 2.7). Fosmids were pooled in equimolar amounts. The fosmid pool was then sequenced with a Roche FLX sequencer (Basel, Switzerland) by M. Schilhabel at the *Institut für klinische Molekularbiologie* at the *Christian Albrechts Universität zu Kiel*.

2.11.3 Analysis of DNA and protein sequences

The analysis of sequence data was performed applying the following software.

ARB® (Ludwig et al. 2004)
BioEdit
Chromas Lite Technelysium Pty Ltd
Clone Manager Suite 7 Scientific & Educational Software
Staden Package containing Pregap4 and Gap4

2.12 Construction of fosmid libraries

Fosmid libraries were constructed with the CopyControl™ Fosmid Library Production Kit from Epicentre (Madison, WI, USA) according to the instruction manual.

2.12.1 End-Repair

After the isolation (2.5.1) and analysis of the metagenomic DNA for its purity, integrity and quantity (2.6, 2.7), its ends were adjusted to blunt ends. Therefore the following reaction was performed:

End-Repair buffer (10x)	4 µL
dNTPs (2.5 mM)	4 µL
ATP (10 mM)	4 µL
DNA	up to 10 µg
End-Repair Enzyme Mix	2 µL
H_2O *ad* 40 µL	

The reaction mixture was incubated for 45 min at room temperature and then inactivated in 10 min at 70°C.

2.12.2 Ligation

The end-repaired DNA (2.12.1) and the vector pCC1FOS (2.1) were ligated at a molar ratio of 1:10. The following reaction was incubated for 2 h at room temperature before it was inactivated at 70°C for 10 min.

Fast-Link Ligation buffer (10x)	0.5 µL
ATP (10 mM)	0.5 µL
pCC1FOS (0.5 µg/µL)	0.5 µL
Insert DNA	0.1-3 µL*
Fast-Link ligase	0.5 µL
H_2O *ad* 5 µL	

* E.g. 0.5 µL (0.25 µg) vector would be ligated with 0.125 µg DNA with a size of approximately 40 kb.

2.12.3 Packaging of fosmid clones

To package the ligation (2.12.2) in phage particles, one tube of MaxPlax Lambda Packaging Extracts was thawed and one half added to the ligation, the remaining packaging extract was again stored at -70°C. The reaction was then incubated for 90 min at 30°C before the other half of the packaging extract was added to the reaction that was then again incubated at 30°C for 90 min. Dilution buffer was added to a final volume of 1 mL. Then 25 µL chloroform were added and the reaction was mixed briefly and stored at 4°C.

Dilution buffer

Tris	10 mM
NaCl	100 mM
$MgCl_2$	10 mM

The pH was adjusted to 8.3.

2.12.4 Preparation of phage competent cells

To prepare phage competent cells, 50 mL LB (2.2.2; + 10 mM $MgSO_4$) were inoculated with an overnight culture of EPI300 (2.1) cells and incubated at 37°C until an OD_{600} of 0.8-1.0 was reached. The cells were stored at 4°C until further use.

2.12.5 Transduction

For the transduction, 10 µL packaged fosmids (2.12.3) were added to 100 µL competent cells (2.12.4) and incubated for 20 min at 37°C. The infected cells were spread on LB plates (2.2.2) containing IPTG, X-Gal and 12.5 µg/mL chloramphenicol (2.2.1).

Several individual clones were analysed for insert frequency and size by induction (2.12.6), fosmid preparation (2.5.4), restriction analysis (2.9.1) and agarose gel electrophoresis (2.7).

2.12.6 Induction

For the analysis of fosmid clones, their induction from a single copy vector to a high copy vector is possible. To induce the increase of the copy number, LB medium (2.2.2) with 12.5 µg/mL chloramphenicol was inoculated with 10% (vol/vol) of an overnight culture and supplied with 1‰ (vol/vol) induction solution (supplied within the kit). The culture was then incubated for 5 h at 37°C and the cells subsequently harvested for further analysis.

2.12.7 Storage of metagenomic libraries

The fosmid clones were picked in wells of microtiter plates which were filled with 150 µL LB + 12.5 µg/mL chloramphenicol (2.2.1, 2.2.2) The plates were incubated overnight at 37°C and then supplied with 33% (vol/vol) glycerol (final volume) and stored at -70°C.

2.13 Protein biochemical methods

2.13.1 Cell lysis

For the investigation of proteins produced by bacteria, cell lysis was performed using an ultrasonicator UP200S with microtip S2 (Hielscher Ultrasonics, Teltow, Germany).

Therefore the bacteria from the respective cultures were sedimented at 4,500 *g* at 4°C for 10 min. The supernatant was discarded and the pellet suspended in 1/10 culture vol. appropriate buffer, e.g. 50 mM Tris, pH 8.0. The samples were ultrasonicated on ice with an amplitude of 50% and a cycle of 0.5 until the sample was cleared. Cell debris was sedimented at 4,500 *g* and 4°C for 5 min. The crude cell extracts were stored at 4°C until further use.

2.13.2 Quantification of proteins

For the quantification of proteins, the method of Bradford was applied (Bradford 1976). Therefore 10 µL protein solution were added to 1 mL Bradford solution and incubated for 15 min in the dark. Afterwards the protein concentration was quantified at 595 nm in dependence of a calibration line for which bovine serum albumin (BSA) was used as a standard.

Bradford solution

Coomassie Brilliant Blue G250	100 mg
Ethanol (96%, vol/vol)	50 mL
H_3PO_4 (85%, w/vol)	100 mL
H_2O *ad* 1000 mL	

The solution was filtered and stored at 4°C and protected from light.

2.13.3 SDS-Polyacrylamide gel electrophoresis (PAGE)

For the determination of the molecular weight of proteins, SDS-PAGE was applied.

To prepare the gel, the two glass plates were adjusted in the respective gadget (BIORAD, Hercules, CA, USA). The separation gel was prepared, transferred in the space between the glass plates and allowed to harden for at least 1 h. Then

the stacking gel was prepared and transferred onto the separation gel and allowed to harden for 1 h after the comb was plunged into the gel.

Ingredients for two separation gels (12%)

Acrylamide (40%, w/vol)	3 mL
Separation gel buffer	2.5 mL
H_2O	4.5 mL
TEMED	9 µL
APS (10%, w/vol)	45 µL

Ingredients for two stacking gels (6%)

Acrylamide (40%, w/vol)	0.6 mL
Stacking gel buffer	0.96 mL
H_2O	2.44 mL
TEMED	2 µL
APS (10%, w/vol)	18 µL

Separation gel buffer

Tris	1.5 M
SDS	0.4% (w/vol)

The pH was adjusted to 8.8.

Stacking gel buffer

Tris	500 mM
SDS	0.4% (w/vol)

The pH was adjusted to 6.8.

The samples were boiled for 10 min with 1 vol. sample buffer and then applied onto the gel. Electrophoresis was performed in a Mini-PROTEAN® 3 Cell from BIORAD (Hercules, CA, USA) filled with electrophoresis buffer at 20 mA until the samples reached the separation gel and then at 30 mA. After separation, the gels were stained with Coomassie Blue solution for at least 2 h and destained with 20% (vol/vol) acetic acid. The molecular weight of the proteins was determined by comparing it with the Protein Molecular Weight Marker SM0431 or the Page Ruler Protein ladder SM0661 (Fermentas, St. Leon-Rot, Germany).

Electrophoresis buffer (10x)

Tris	30.3 g
Glycine	144.1 g
SDS	10 g
H_2O *ad* 1000 mL	

The pH was adjusted to 8.4.

Sample buffer

Glycerol	50% (vol/vol)
DTT	0.154 g
SDS	4% (w/vol)
Bromophenol blue	2 mg
Tris	150 mM
EDTA	1 mM
NaCl	30 mM
NaN_3	1 mM
H_2O *ad* 10 mL	

Coomassie Blue solution

Coomassie Brilliant Blue	1 g
Methanol	400 mL
Acetic acid	100 mL
H_2O *ad* 1000 mL	

The solution was filtered and stored at 4°C and protected from light.

2.13.4 Purification of recombinant proteins

To purify recombinant proteins, the respective genes were amplified (2.8) and ligated into an expression vector (2.1). The constructs were transformed (2.10.2) in an *E. coli* strain for overexpression (2.1). The enzyme production was induced with IPTG before the enzymes were purified with the respective columns (2.13.4.1 and 2.13.4.2).

2.13.4.1 pET vectors/His-tag affinity columns

Proteins containing an His-tag sequence were purified with Protino® Ni-TED 2000 packed columns (Macherey-Nagel, Düren, Germany) as described in the manufacturer's protocol. The required buffers were supplied within the kit.

For the induction of the production of proteins encoded by genes ligated in pET vectors (2.1), bacterial cells containing the respective construct were grown

overnight in LB medium (2.2.2) with the appropriate antibiotic. Then 100 mL LB (+ antibiotic) were inoculated 1‰ (vol/vol) with the overnight culture and grown until an OD_{600} of 0.8 to 1.0. Overexpression was induced with 100 µM IPTG and the culture was then incubated for 3 h at an appropriate temperature. The cells were harvested by centrifugation, resuspended in LEW buffer (2 to 5 mL for 1 g wet cells) and lysed by sonication (2.13.1). Cell debris was eliminated by centrifugation (4°C, 10 min, 4,500 *g*). The column was equilibrated with 4 mL LEW buffer. Then the lysate was transferred onto the column and the column was allowed to empty by gravity flow. The column was washed by twice applying 4 mL LEW buffer before the elution of the recombinant protein was performed by applying 3 times 3 mL elution buffer. Molecular weight, purity and activity of the protein were analysed by SDS-PAGE (213.3) and DNSA assay (2.14.2).

2.13.4.2 pMAL-c2/amylose resin affinity columns

Recombinant proteins containing a maltose binding protein domain (MBP) were purified according to the instruction manual from New England Biolabs (Beverly, MA, USA) for pMAL Protein Fusion & Purification System.

For the overexpression of genes that were ligated into pMAL-c2, the respective clone was grown overnight in rich medium (2.2.3) containing ampicillin (2.2.1). A volume of 100 mL was inoculated 1‰ (vol/vol) with the overnight culture and grown to an OD_{600} of 0.5. Overexpression was induced with the addition of 300 µM IPTG. The culture was then incubated for 3 h at an appropriate temperature. The induced cells from 100 mL culture were sedimented by centrifugation (4°C, 10 min, 4,500 *g*) and suspended in 5 mL column buffer. Cell lysis was performed by sonication (2.13.1). Cell debris was eliminated by centrifugation (4,500 *g*, 4°C, 10 min) and the supernatant was diluted 1:6 with column buffer. The column was prepared as follows: amylase resin (3 mL) was poured into the column and then washed with 5 vol. column buffer. Afterwards the diluted crude extract was transferred onto the column before the column was washed again with 12 vol. column buffer. The target protein was eluted in 10 mL column buffer + 10 mM maltose. The purification steps were analysed by SDS-PAGE (2.13.3) and DNSA assay (2.14.2).

Column buffer

Tris-HCl	20 mM
NaCl	200 mM
EDTA	1 mM

The pH was adjusted to 7.4.

2.14 Assays for the detection and quantification of cellulolytic activity

For the detection of cellulolytic clones, the congo red plate assay (2.14.1) was applied. The quantification of activity was performed with the DNSA assay (2.14.2).

2.14.1 Congo red plate assay

For the detection of cellulolytic activity, strains and metagenomic clones were streaked or stamped on appropriate agar plates (2.2) containing 0.2% (w/vol) CMC. The plates were incubated for 2 to 7 days under the respective growth conditions (2.4) and then overlaid with 0.2% (vol/vol) congo red. The plates were incubated for 30 min before the congo red solution was discarded and the plates were washed with 1 M NaCl twice for 30 min. Positive clones exhibited an orange halo on the red stained plates.

2.14.2 3,5-dinitrosalicylic acid (DNSA) assay

To quantify cellulolytic activity, crude cell extract or purified enzymes were used in the DNSA assay. The activity was determined by the amount of reducing sugar released from cellulose as this sugar reduces 3,5-dinitrosalicylic acid (DNSA) to 3-amino-5-nitrosalicylic acid which can be quantified at 546 nm.

2.14.2.1 Standard DNSA assay

The standard assay mixture contained

Crude cell extract/recombinant protein	100 µL
CMC (2%, vol/vol)	250 µL
McIlIvaine buffer	150 µL

This reaction mixture was incubated for 30 min to 2 h at an appropriate temperature. Then 750 µL DNSA reagent were added and the sample incubated for 15 min at 95°C. The reducing sugar ends were quantified at 546 nm in a SmartSpec™ Plus spectrophotometer (Bio-Rad, Hercules, CA, USA) in comparison with a calibration line for which glucose was used as a standard.

McIllvaine buffer

Na_2HPO_4	0.2 M

The pH of 6.5 was adjusted with 0.1 M citric acid at 65°C.

DNSA reagent

3,5-dinitrosalicylic acid	10 g
NaOH	10 g
K-Na-Tartrat	200 g
Na_2SO_3	0.5 g
Phenol	2 g
H_2O *ad* 1000 mL	

The solution was stored at 4°C and protected from light.

2.14.2.2 Measurement of cellulolytic activity in the presence of ILs

Cellulolytic activity in the presence of ILs was routinely measured with 30% (vol/vol) IL. The same reaction mixture as described in 2.14.2 was prepared, but McIllvaine buffer was replaced with one of the ILs. For the measurement of activity in the presence of varying IL volumes these were adjusted. If the IL content was higher than 30%, the content of CMC was reduced.

The ILs used in this study were: 1-butyl-3-methylimidazolium chloride ([BMIM]Cl), 1-butyl-3-methylimidazolium trifluoromethanesulfonate ([BMIM][OTF]), 1-butyl-2,3-dimethylimidazolium chloride ([BMMIM]Cl), 1-ethyl-3-methylimidazolium trifluoroacetate ([EMIM][ATF]), 1-butyl-1-methylpyrrolidinium trifluoromethanesulfonate ([BMPL][OTF]) and 1-ethyl-3-methylimidazolium trifluoromethanesulfonate ([EMIM][OTF]) (Fig. 2.1).

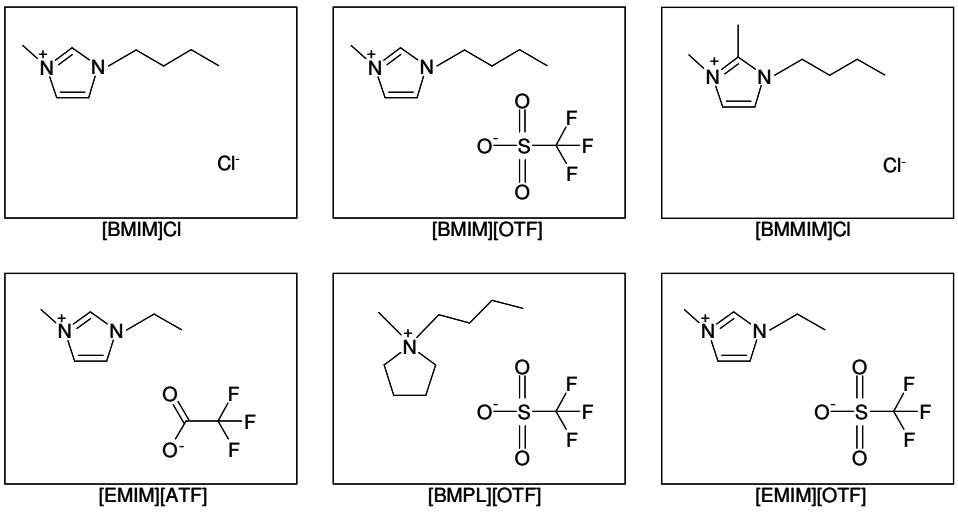

Figure 2.1: Ionic liquids used in this study

The ILs were supplied by Merck (Darmstadt, Germany) and were all liquid at room temperature with the exception of [BMIM]Cl and [BMMIM]Cl. The melting points of these two ILs were lowered with the addition of 8.6% (vol/w) H_2O for [BMIM]Cl and 13.6% (vol/w) H_2O for [BMMIM]Cl, respectively, and therewith liquefied.

2.14.2.3 Measurement of the long term stability of cellulases

To test the enzymes' long term stability under specific conditions, 100 µL of an enzyme solution were added to 150 µL McIllvaine buffer or IL or any other desired reagent and incubated for the respective time at the respective temperature. Then the substrate was added and the assay was continued as described in 2.14.2. For measuring the stability in the presence of ILs, the pre-incubation therefore was performed in 60% and the activity measurement in 30% IL (vol/vol). For the measurement of activity and stability in varying NaCl conditions, McIllvaine buffer and the substrate were both supplied with NaCl.

2.14.2.4 Measurement of cellulolytic activity against alternative substrates

As alternative substrates, crystalline cellulose and crystalline cellulose that was dissolved in IL and precipitated with water were used, these replaced CMC in the standard assay (2.14.2).

Crystalline cellulose that has been dissolved in 1-butyl-3-methylimidazolium acetate [BMIM]Ac or 1-ethyl-3-methylimidazolium acetate [EMIM]Ac and

afterwards precipitated with water was received from S. Endisch (*Fraunhofer ICT*, Pfinztal, Germany). The final concentration of IL in the assays was 4.32% (vol/vol).

3 Results

Cellulases can be used for a wide variety of biotechnological applications, e.g. in paper and textile industries. Of main interest is that cellulases can be used for the hydrolysis of cellulosic plant material for the further processing to valuable molecules like ethanol. The major problem of enzymatic cellulose hydrolysis is the insolubility of crystalline cellulose in water. Cellulases, however need an aquatic medium for industrially relevant conversion rates. As some ionic liquids (ILs) can dissolve cellulose, they could be used as an alternative reaction medium if cellulases were active and stable in their presence.
The aim of this study was to discover novel cellulases that are active and stable in the presence of ILs. Therefore a culture-independent method, metagenomics, was applied. Different habitats were chosen for the construction of metagenomic libraries that were subsequently screened for clones with cellulolytic activity. Positive clones were analysed with respect to their activity in the presence of ILs and potentially related characteristics. As the three most IL-tolerant enzymes were moderately thermostable, they were benchmarked with two further thermostable enzymes; CelA from *Thermotoga maritima* (Liebl et al. 1996) and a metagenomic cellulase derived from the Avachinsky Crater in Kamchatka designated as Cel5K.

3.1 Evaluation of habitats

Prior to the construction of metagenomic libraries, different habitats supposed to comprise a wide variety of cellulolytic microorganisms were evaluated on activity and on biodiversity level. The evaluated habitats were faeces of an Asian elephant (2.3.1.1), a biogas plant sample (2.3.1.4) and enrichment cultures (2.4.1) inoculated with elephant faeces, koala faeces (2.3.1.2), the biogas plant sample and with an extract of the shipworm *Teredo navalis* (2.3.1.3).
All enrichment cultures were grown in MSM medium (2.2.4) with a cellulosic substrate as sole carbon source. The enrichments inoculated with elephant faeces, the biogas plant sample and *Teredo navalis* were supplied with 0.5% (w/vol) carboxymethylcellulose (CMC) while the culture inoculated with koala

faeces was supplied with 1% (w/vol) cellulosic filter paper. Except the culture inoculated with *Teredo navalis*, which was incubated aerobically at 22°C, all cultures were incubated at 37°C under microaerobic conditions.

3.1.1 Cellulolytic activity of the consortia inhabiting the evaluated habitats

The cellulolytic activity of the microbes inhabiting the different habitats, elephant faeces, the biogas plant sample and enrichment cultures inoculated with elephant faeces, koala faeces, the biogas plant sample and with an extract of *Teredo navalis*, was investigated using the DNSA method with CMC as a substrate (2.14.2). As shown in Fig. 3.1, the crude extracts of the directly investigated habitats, elephant faeces and the biogas plant sample, were by far more active than each of the enrichment cultures, which were evaluated after 4 months. The cellulolytic activity of the metaproteome of the elephant faeces was 165 mU/mg and that of the metaproteome of the biogas plant sample 60 mU/mg. The activity of the cell extracts was reduced significantly by the addition of 30% (vol/vol) IL; the activity of the biogas plant sample metaproteome was reduced to about 50% relative activity and that of the elephant faeces to about 10% relative activity (Fig. 3.1).

Among the enrichment cultures, the metaproteome of the *Teredo* culture was, with 2 mU/mg, the most active, though none of the metaproteomes of the enrichment cultures showed high activity, and in the presence of 30% (vol/vol) IL there was almost no activity left (Fig. 3.1).

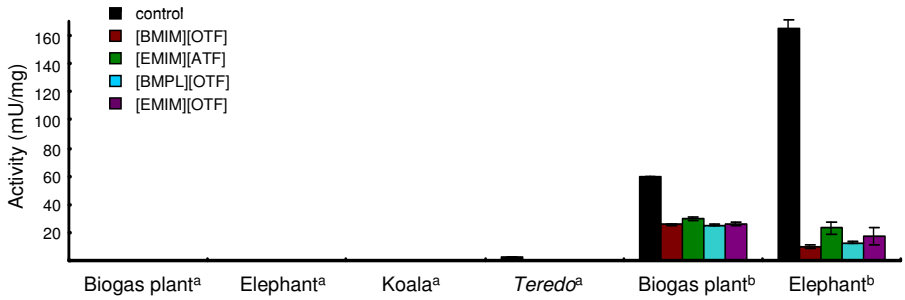

Figure 3.1: Cellulolytic activity of the metaproteomes of the investigated habitats

In the activity measurements with ILs the concentration of IL was 30% (vol/vol). [a]: habitat was an enrichment culture, [b]: habitat was directly investigated. Activities were measured at 37°C.

The biogas plant sample, elephant faeces and the enrichment culture inoculated with an extract of *Teredo navalis* were chosen for further investigation.

3.1.2 Microbial diversity of the investigated habitats

The microbial diversity of the elephant faeces (2.3.1.1) and the enrichment culture inoculated with an extract of *Teredo navalis* (2.3.1.3) were investigated via 16S rRNA gene analysis. Therefore metagenomic DNA was extracted (2.5.1) and a PCR was performed using the 16S PCR primers 616V and 1492R (2.1). The resulting PCR fragments were ligated into pDrive (2.9.3.1), transformed (2.10.2) in *E. coli* DH5α (2.1) and sequenced (2.11.1).

3.1.2.1 Elephant faeces

For the phylogenetic characterisation of the microbiota inhabiting the faeces of the Asian elephant, 198 16S rRNA gene clones were analysed. The microbiota of the elephant faeces was highly diverse and comprised organisms of all major phyla. The closest related sequences of the largest fraction of the amplified 16S rRNA genes belonged to bacteria of the phylum *Proteobacteria*, which were dominated by *γ-Proteobacteria* (Fig. 3.2). Additionally *Firmicutes* and *Bacteroidetes* were highly abundant. *Actinobacteria*, *Lentisphaera*, *Verrucomicrobia* and *Spirochaetes* have been detected furthermore (Fig. 3.2).

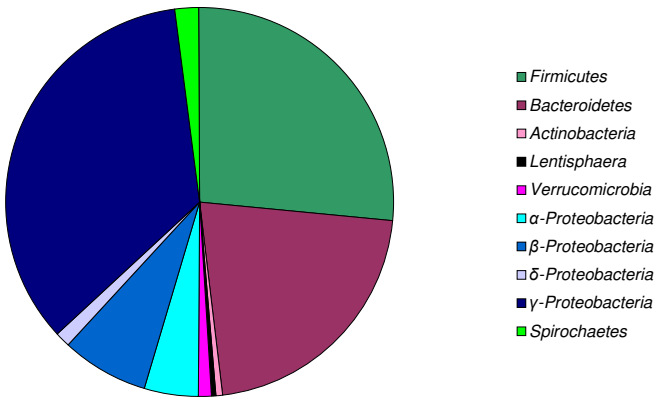

Figure 3.2: Frequency of phyla in the faeces of the Asian elephant (2.3.1.1) based on 16S rRNA gene analysis

Additionally many of the 16S rRNA genes, especially those with highest identity to sequences of bacteria belonging to *Firmicutes*, clustered with 16S rRNA sequences from so far uncultivated organisms or were closely related with sequences of microbes that are known to be capable of hydrolysing cellulose like *Ruminococcus albus* or *Butyrivibrio fibrisolvens* (Fig. 3.3). Many sequences were also in high distance to their closest related known sequence (Fig. 3.3).

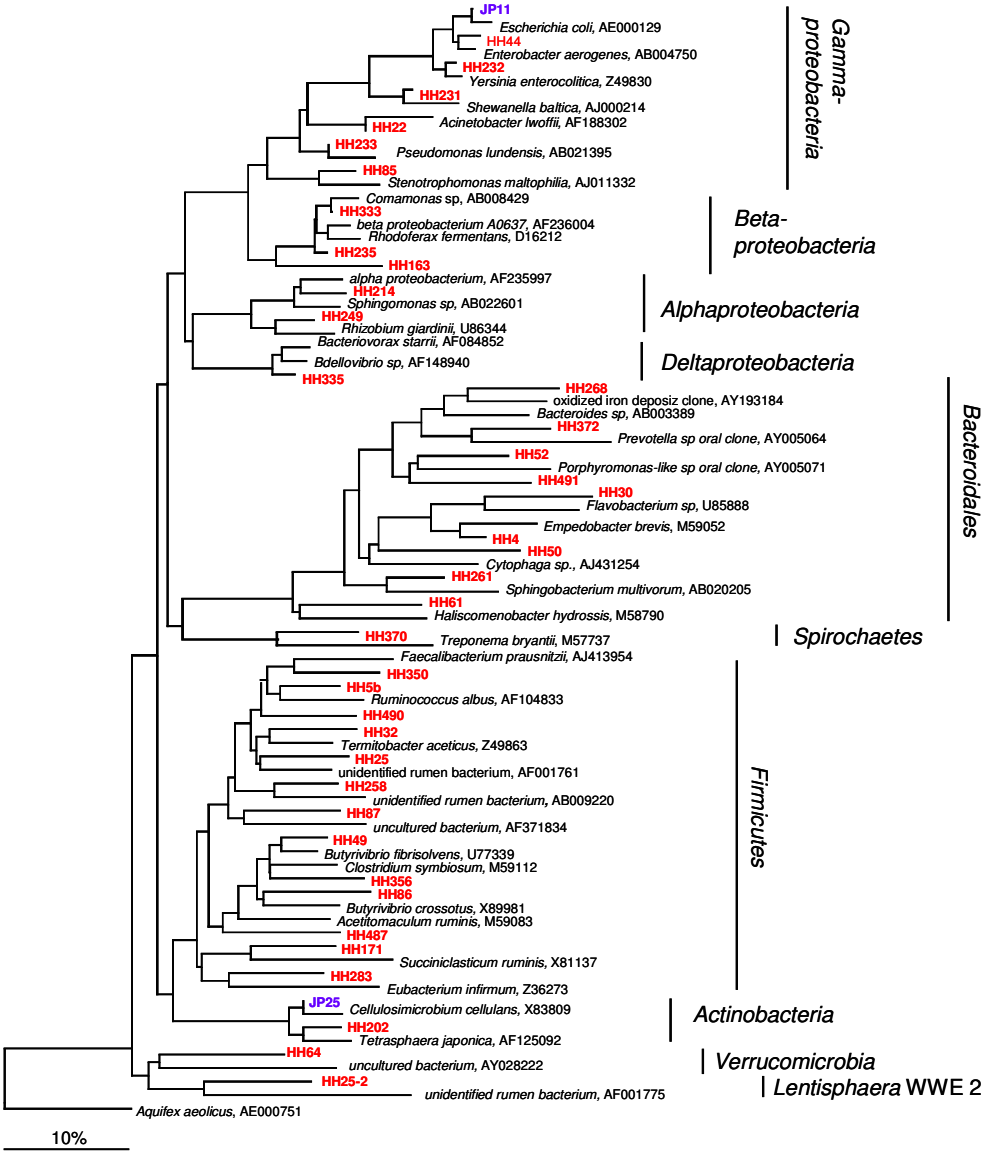

Figure 3.3: Phylogenetic tree based on 16S rRNA gene sequences derived from the faeces of the Asian elephant (2.3.1.1)

Sequences marked in red with the prefix HH derived directly from the faeces, purple sequences with the prefix JP derived from an enrichment culture in MSM (2.2.4) and cellulosic filter paper that was inoculated with elephant faeces and incubated at 37°C (Pottkämper 2009), sequences marked in black indicate known GenBank entries. The tree was constructed using ARB software and the Maximum Likelihood method (2.11.3).

3.1.2.2 *Teredo navalis* enrichment culture

The analysis of 20 16S rRNA gene sequence clones derived from the *Teredo* culture revealed that the community mainly consisted of *Proteobacteria* that were dominated by *γ-Proteobacteria* (Fig. 3.4). Many of the sequences showed highest identity to the 16S rRNA gene of *Acinetobacter* species (Fig. 3.5). Additionally *Firmicutes*, *Actinobacteria* and *Bacteroidetes* were detected in this enrichment culture (Fig. 3.4).

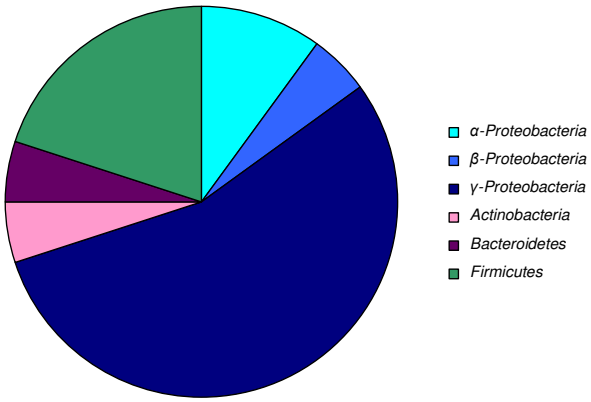

Figure 3.4: Frequency of phyla of the microbiota of the enrichment culture inoculated with an extract of *Teredo navalis* (2.3.1.3) based on 16S rRNA gene sequence analysis

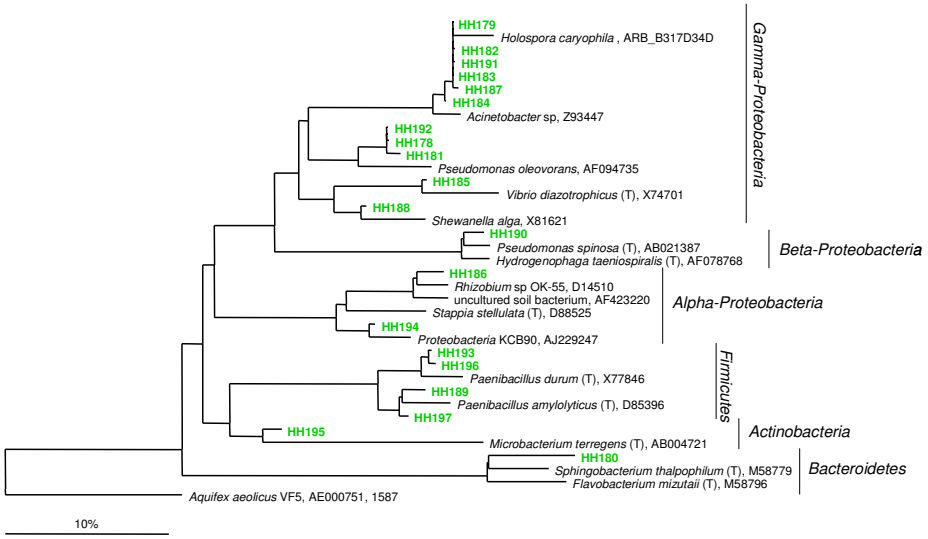

Figure 3.5: Phylogenetic tree based on 16S rRNA gene sequences derived from the enrichment culture inoculated with an extract of *Teredo navalis* (2.3.1.3)

Sequences marked in green derived from the enrichment culture, sequences marked in black indicate known GenBank entries. The tree was constructed using ARB software and the Maximum Likelihood method (2.11.3).

3.2 Construction of metagenomic libraries

In this work a metagenomic library was constructed (2.12) with DNA isolated (2.5.1) from the enrichment culture inoculated with an extract of *Teredo navalis* (2.4.1). This library comprised 3,600 clones. To determine the insert frequency and size, 10 fosmids were purified (2.5.4) and an analytical restriction was performed using *Bam*HI and *Eco*RI (2.9.1). The insert frequency was 100% and the average insert size was 27.5 kb (data not shown).

3.3 Screening of metagenomic libraries

An overview of the libraries investigated in this study is given in Table 3.1. In this work, the libraries constructed with DNA from elephant faeces and from the enrichment culture inoculated with an extract of *Teredo navalis* with a total of 23,600 clones were screened for cellulolytic activity. Therefore the libraries were stamped on LB (2.2.2) agar plates containing CMC. After an incubation period of 2 to 4 days, the plates were stained with congo red dye (2.14.1). The plates were then washed with 1 M NaCl and an orange halo occurred around cellulolytic clones (Fig. 3.6).

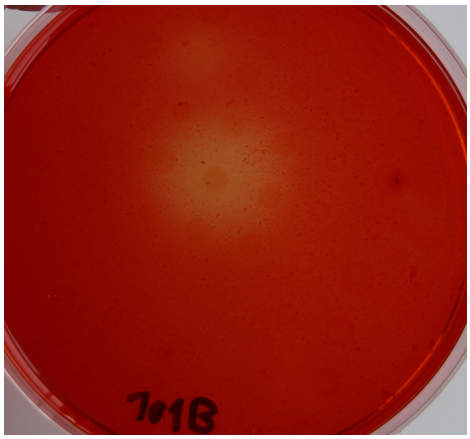

Figure 3.6: Clone with cellulolytic activity on a CMC agar plate stained with congo red

By screening the metagenomic libraries, 12 positive clones were detected (Table 3.1) (the library constructed with DNA isolated from the biogas plant was

screened for cellulolytic activity by D. Meske, whereby two clones where chosen for further investigation in this study).

Table 3.1: Metagenomic libraries investigated in this study and the discovered cellulolytic fosmid clones

Library	No. of clones screened	Average insert size (kb)	No. of cellulolytic clones detected	No. of IL-tolerant cellulolytic clones obtained*
Teredo enrichment	3,600	27.5	1	0
Elephant faeces	20,000	37.5	11	7
Biogas plant	9,600	23.5	2	2

* IL-tolerant means that the clone showed activity and stability in at least one of the used ILs

3.4 Investigation of putative positive clones on activity level

All quantitative activity measurements were performed with the DNSA assay (2.14.2).

3.4.1 Activity of the cellulolytic clones in the presence of 30% (vol/vol) IL

As the aim of this study was to identify cellulases that are active and stable in the presence of ILs, the discovered cellulolytic fosmid clones derived from the libraries constructed with DNA from elephant faeces, the biogas plant and the *Teredo* culture were assayed in the presence of 30% (vol/vol) of the different ILs (2.14.2.2). Therefore crude extracts (2.13.1) of induced (2.12.6) fosmid clones were investigated with the DNSA assay (2.14.2). With the exception of the clone derived from the library constructed with DNA isolated from the *Teredo* enrichment culture, all clones showed high activity in the control assays without IL (Table 3.2). In most cases the activity in the presence of ILs was significantly lower than in the assays without IL. Especially the clones 89E1 and 94A1 lost activity drastically with the addition of 30% (vol/vol) IL. Nevertheless, many of the clones still revealed high activity (Table 3.2). Especially the activity of the clones derived from the library derived from the biogas plant sample and of several clones derived from the library constructed with DNA from elephant faeces (including 82F10, pFosCelA84, 89C12, 101D9 and 179F2) was not reduced severely. It was also shown that the ILs [BMIM]Cl and [BMMIM]Cl

generally reduced the enzymes' activity stronger than the other investigated ILs (Table 3.2).

Table 3.2: The activity of crude extracts of the discovered fosmid clones in the presence of 30% (vol/vol) of the different ILs

Clone	Metagenomic library	Activity (mU/mg)						
		control	[BMIM] Cl	[BMIM] [OTF]	[BMMIM] Cl	[EMIM] [ATF]	[BMPL] [OTF]	[EMIM] [OTF]
pFosCelA2	Biogas plant	507±6	192±5	461±39	181±10	506±15	470±10	516±26
pFosCelA3	Biogas plant	76±2	49±3	70±17	55±2	75±1	77±2	69±1
pFosCelA31	*Teredo* culture	3±0	1±0	1±0	1±0	0	0	0
68D12	Elephant faeces	70±5	2±0	15±1	0	21±3	3±0	14±4
82F10	Elephant faeces	74±9	2±0	24±2	0	36±3	4±0	22±5
83D12	Elephant faeces	108±13	3±0	19±5	0	26±9	30±2	8±2
pFosCelA84	Elephant faeces	83±5	5±1	4±1	2±0	33±2	24±2	42±6
89C12	Elephant faeces	188±6	8±2	111±2	6±3	109±2	37±2	55±3
89E1	Elephant faeces	37±4	0	1±0	0	7±3	0	0
94A1	Elephant faeces	102±3	0	2±1	0	2±0	0	0
101C1	Elephant faeces	86±6	1±0	37±3	0	44±1	1±1	29±5
101D9	Elephant faeces	95±9	2±0	60±2	0	60±0	19±4	43±9
172E9	Elephant faeces	173±19	0	0	0	29±5	0	33±2
179F2	Elephant faeces	612±6	4±2	239±3	2±0	163±13	0	220±14

The control did not include any IL

Furthermore the cellulolytic activity of the crude extracts of the fosmid clones was assayed after overnight incubation at room temperature in 60% (vol/vol) [BMIM][OTF], [EMIM][ATF], [BMPL][OTF] and [EMIM][OTF]. Thereby the activity was measured at 37°C after the addition of the substrate whereby the IL concentration was reduced to 30% (vol/vol). Under these conditions, many clones lost a high proportion of their activity (Table 3.3). E.g. the clones 101C1 and 172E9, which were highly active in the presence of 30% (vol/vol) IL, lost almost their complete activity after overnight incubation in 60% (vol/vol) IL. Nevertheless some of the clones, e.g. pFosCelA2, pFosCelA3, 82F10, pFosCelA84, 89C12, 101D9 and 179F2, still showed high activity in the presence of at least one of the tested ILs (Table 3.3). The clones derived from the library constructed with DNA from the biogas plant (pFosCelA2 and

pFosCelA3) and pFosCelA84 from the library constructed with DNA from elephant faeces were chosen for further characterisation as these were the clones that showed the highest activity and stability in the presence of a broad range of ILs (Table 3.2, 3.3).

Table 3.3: The activity of crude extracts of the discovered fosmid clones after overnight incubation in the presence of 60% (vol/vol) of different ILs at room temperature, the activity was then assayed in the presence of 30% (vol/vol) IL at 37°C after the addition of the substrate CMC

Clone	Metagenomic library	Activity (mU/mg)				
		control	[BMIM][OTF]	[EMIM][ATF]	[BMPL][OTF]	[EMIM][OTF]
pFosCelA2	Biogas plant	157±21	0	5±2	94±1	119±19
pFosCelA3	Biogas plant	100±1	3±0	50±6	46±1	49±5
pFosCelA31	*Teredo* culture	1±0	0	0	0	0
68D12	Elephant faeces	94±8	4±1	0	0	16±0
82F10	Elephant faeces	98±1	10±1	0	1±0	21±1
83D12	Elephant faeces	101±0	7±1	1±0	9±0	5±1
pFosCelA84	Elephant faeces	145±2	28±1	23±3	0	111±6
89C12	Elephant faeces	136±8	3±1	0	0	28±1
89E1	Elephant faeces	33±2	0	0	0	0
94A1	Elephant faeces	8±0	0	0	0	0
101C1	Elephant faeces	8±1	0	0	0	1±0
101D9	Elephant faeces	113±15	21±1	0	0	24±1
172E9	Elephant faeces	385±10	0	0	0	0
179F2	Elephant faeces	1162±30	0	0	0	267±49

The control did not include any IL

3.4.2 Investigation of the fosmid clones' activity in different concentrations of ILs

Crude extracts (2.13.1) of the metagenomic fosmid clones carrying pFosCelA2, pFosCelA3 or pFosCelA84 were assayed at 37°C in the presence of 10 to 60% (vol/vol) of the different ILs used in this study. The activity of the crude extract of clone pFosCelA84 was strongly affected by the addition of ILs. The addition of only 10% (vol/vol) [BMIM]Cl or [BMMIM]Cl resulted in a reduction of activity to 50 or 35% relative activity, respectively (Fig. 3.7C). In the presence of 10 to 30% (vol/vol) of the ILs [EMIM][ATF] and [EMIM][OTF] 30-70% relative activity was left. When the ILs [BMIM][OTF] and [BMPL][OTF] were added, the activity

was significantly reduced by 10% (vol/vol) IL (70 and 50% relative activity, respectively), but in the presence of 20 to 60% (vol/vol) IL the crude extract of pFosCelA84 still revealed 10 to 20% relative activity (Fig. 3.7.C).

The cellulolytic activity of the crude extracts of the clones pFosCelA2 and pFosCelA3 was less affected by the addition of ILs (Fig. 3.7.A, B). In the presence of 10 to 20% (vol/vol) IL the activity was only slightly reduced, though that of pFosCelA2 was stronger reduced by [BMIM]Cl, [BMIM][OTF] and [BMMIM]Cl than by the other ILs (Fig. 3.7.A). When 30 to 40% (vol/vol) IL were added, the activity was on average about 40% of the original activity and in the presence of 50 and 60% IL hardly any activity was left (Fig. 3.7.A, B).

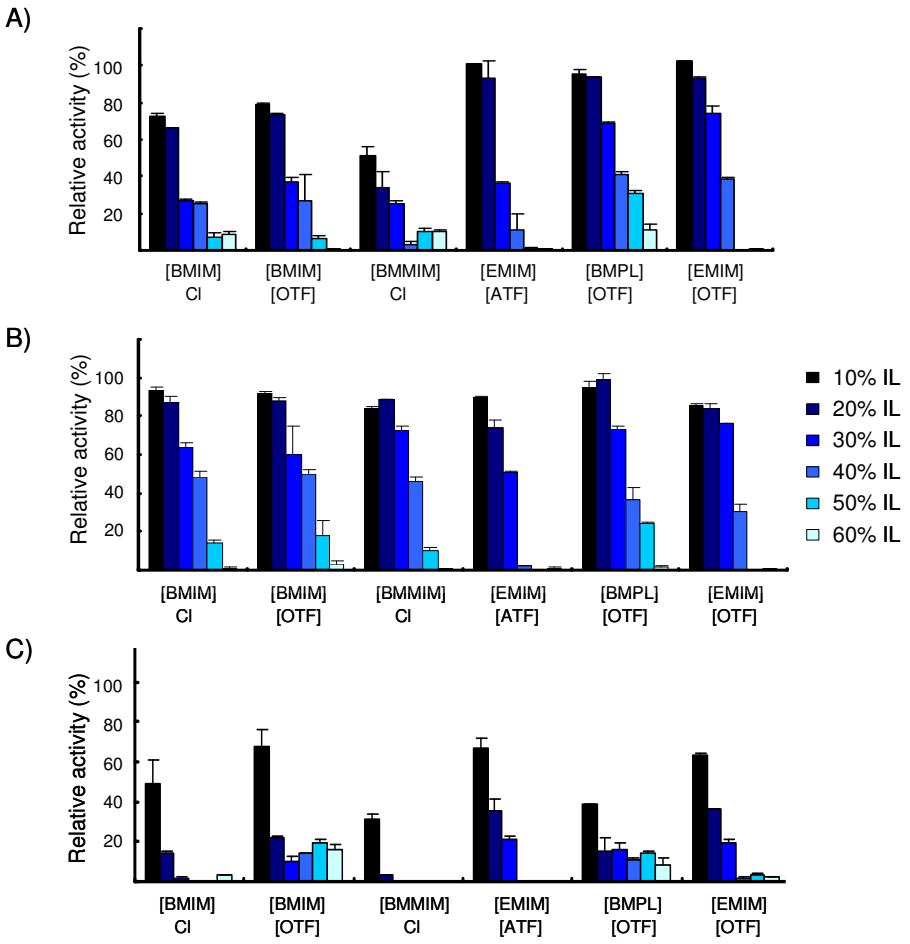

Figure 3.7: Relative activity of the crude extracts of the metagenomic clones in the presence of IL concentrations ranging from 10 to 60% (vol/vol) of different ILs

A) pFosCelA2, B) pFosCelA3, C) pFosCelA84. The activity in the control assay without IL was set as 100%. Activities were measured at 37°C.

3.4.3 Evaluation of the NaCl tolerance of the cellulolytic fosmid clones

To furthermore investigate halotolerance, the activity and the long term stability of the cellulolytic activity of the crude extracts of the fosmid clones pFosCelA2, pFosCelA3 and pFosCelA84 were investigated in the presence of 2 M and 4 M NaCl over a time period longer than 30 days (Fig. 3.8). The incubation was performed at room temperature while the activity was assayed at 37°C with the DNSA assay (2.14.2).

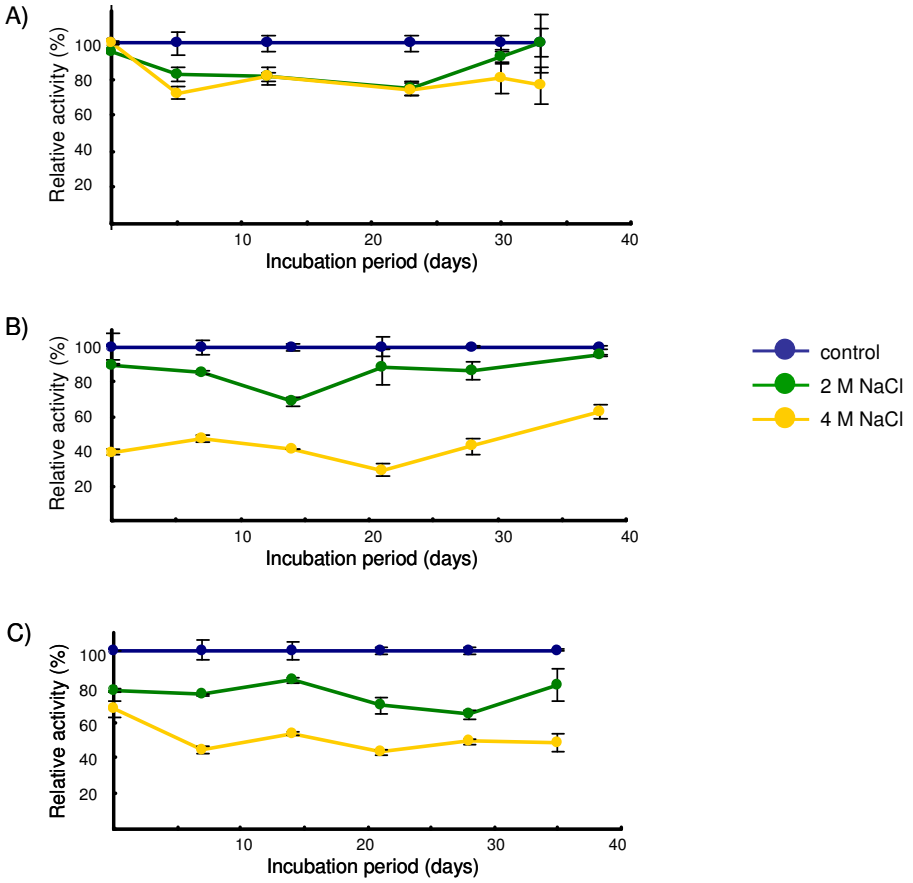

Figure 3.8: Relative activity of crude extracts of the metagenomic fosmid clones in the presence of 2 M and 4 M NaCl over a time period of more than 30 days

A) pFosCelA2, B) pFosCelA3, C) pFosCelA84. The activity of the control assays was set as 100%.

The cellulolytic activity of the crude extracts of all three enzyme clones showed remarkable long term stability in the presence of high salt concentrations. After more than 30 days, the crude extract of clone pFosCelA2 showed 80% relative activity in 4 M NaCl compared to the control (Fig. 3.8.A). After the same time period the crude extracts of the clones pFosCelA3 and pFosCelA84 exhibited

about 50% relative activity in 4 M NaCl. In 2 M NaCl none of the three assayed clones lost more than 20% activity after 30 days compared to the control (Fig. 3.8.A, B, C).

3.4.4 Investigation of the enzymes' tolerance against EDTA, ethanol, glycerol and Tween20

The activity of the crude extracts of the fosmid clones pFosCelA2, pFosCelA3 and pFosCelA84 against CMC was assayed in the presence of some industrially relevant additives like ethanol and EDTA. The addition of 5% (vol/vol) glycerol and 50 mM EDTA did not affect the cellulolytic activity of any of the crude extracts significantly (Fig. 3.9). The activity of the crude extracts of all clones was affected by the addition of ethanol, though 10% (vol/vol) did not affect the cellulolytic activity of the protein extract of clone pFosCelA84. Under the same conditions, the activity of the crude extract of clone pFosCelA2 was reduced to 65% and that of the crude extract of clone pFosCelA3 to 80% relative activity. With the addition of 20% (vol/vol) ethanol the cellulolytic activity was reduced to 40% (pFosCelA2), 65% (pFosCelA3) or 50% (pFosCelA84) relative activity, respectively (Fig. 3.9). In the presence of 20% (vol/vol) Tween20, the activity of the crude extract of clone pFosCelA2 was not affected, while that of the crude extract of clone pFosCelA3 was reduced to 50% and the crude extract of clone pFosCelA84 showed more than 130% relative activity (Fig. 3.9).

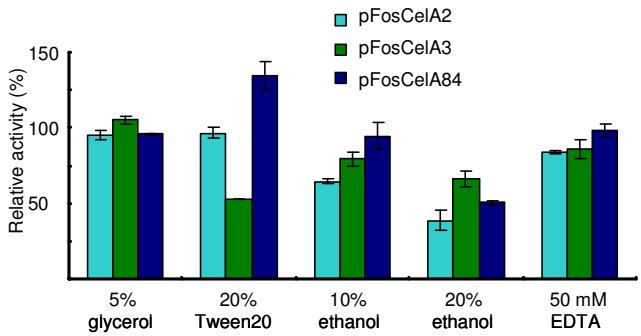

Figure 3.9: Relative cellulolytic activity of crude extracts of the clones pFosCelA2, pFosCelA3 and pFosCelA84 in the presence of 5% (vol/vol) glycerol, 20% (vol/vol) Tween20, 10 and 20% (vol/vol) ethanol and 50 mM EDTA

The activity in the control assays was set as 100%.

3.5 Sequence analysis of fosmid clones

To identify the genes conferring cellulolytic activity to the respective fosmid clones, two different methods were applied. These were subcloning (3.5.1) and pyrosequencing (3.5.2).

3.5.1 Subcloning and primer walking

To obtain the sequence of the gene conferring cellulolytic activity to the fosmid clone pFosCelA31 derived from the library constructed with DNA from the enrichment culture inoculated with an extract of *Teredo navalis*, the fosmid was subcloned. Therefore the fosmid was isolated (2.5.4) and digested with different restriction endonucleases (2.9.1). *Hind*III and *Eco*RI, which produced fragments with a length of 1 to 10 kb, were chosen for a preparative digestion (2.9.1). The obtained fragments were purified (2.5.6) and ligated into pBluescript SKII+ (2.1) that was previously linearised (2.9.1) with the same restriction enzyme and subsequently dephosphorylated (2.9.2). Then the ligation was transformed (2.10.2) in *E. coli* DH5α (2.1). The resulting clones were investigated for cellulolytic activity with the congo red plate assay (2.14.1). Via sequencing (2.11.1) with the primers M13 -20 and M13 rev (2.1) and subsequent primer walking the sequence of the complete gene was obtained; the cellulase gene was designated as *celA31* (Acc. No. JF826523) and the respective enzyme showed highest identity (64%) to a β-1,4-cellobiosidase from *Klebsiella pneumoniae* (77% similarity). The protein consisted of 820 aa and belonged to glycoside hydrolase (GH) family 9. It additionally included an Ig-like domain and a carbohydrate binding module (CBM) (family 4_9).

3.5.2 Pyrosequencing of fosmids

For a deeper insight into the hydrolytic community inhabiting the elephant faeces, several fosmids of cellulolytic fosmid clones were sequenced by pyrosequencing (2.11.2). The fosmids were purified (2.5.4) to a $ratio_{260/280}$ of 1.6-2.0 (2.6) and sequenced as a pool whereby the fosmids were supplied in equimolar amounts. The pool was sequenced by M. Schilhabel (*Institut für klinische Molekularbiologie*, *Christian Albrechts Universität zu Kiel*) and the obtained sequences were related to the respective fosmids with previously

obtained Sanger sequences (2.11.1). Some of the sequenced fragments with putative ORFs are shown in Fig. 3.10 to 3.13.

3.5.2.1 Sequence analysis of pFosCelA84

The cellulase gene *celA84* was identified via pyrosequencing (2.11.2). The sequence of this fosmid was recovered completely (33,820 bp insert). Almost all ORFs on this fragment identified with CloneManager (2.11.3) that coded for proteins with significant similarity to previously annotated proteins, coded for proteins with highest identity to *Fibrobacter succinogenes* (Fig. 3.10). One gene encoding a cellulase was discovered. The respective enzyme CelA84 (Acc. No. JF895785) was 506 aa long and showed highest identity to a cellulase from *Fibrobacter succinogenes* (41% identity/55% similarity). The assumed cellulase belonged to GH family 5 and did not possess any additional domain but it contained with a probability of 0.99 a signal peptide sequence (SignalP-HMM). Except *celA84* no further genes for cell wall degrading enzymes or related functions were identified on fosmid pFosCelA84, but a gene coding for a protein that showed 77% identity to an organic solvent tolerance protein from *Fibrobacter succinogenes* (Fig. 3.10).

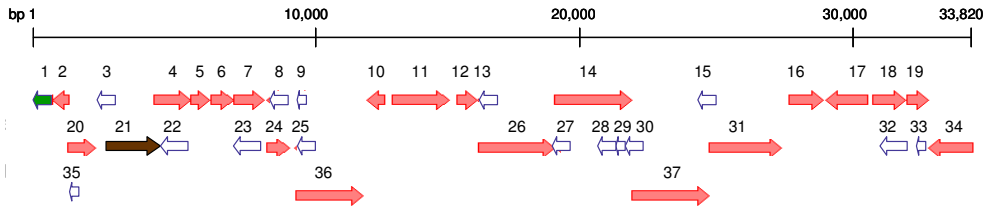

Figure 3.10: Putative ORFs discovered on fosmid pFosCelA84

The phylum of the next known relative (aa identity) is indicated by the colour of the arrow. Green: *Bacteroidetes*, red: *Fibrobacteres*, brown: *Planctobacteria*, white: no significant similarities observed. All putative proteins are listed below in Table 3.4 with their closest relative (% aa identity).

Table 3.4: Putative proteins coded for on fosmid pFosCelA84. Proteins involved in carbohydrate binding and hydrolysis are marked in bold.

ORF No.	aa identity (%) to the next known relative	next known relative
1	65	hypothetical protein [*Bacteroides* sp. 4_3_47FAA] >gb|EET16986.1
2	83	hypothetical protein Fisuc_2766 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76349.1
3		No significant similarities observed
4	89	deoxyguanosinetriphosphate triphosphohydrolase [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76099.1
5	62	pyrimidine 5'-nucleotidase [*Fibrobacter succinogenes* subsp. *succinogenes* S85 >gbIYP_003250582.1
6	82	ABC-3 protein [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76101.1
7	76	hypothetical protein Fisuc_2516 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76102.1
8		No significant similarities observed
9		No significant similarities observed
10	83	non-canonical purine NTP pyrophosphatase, rdgB/HAM1 family [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76155.1
11	70	hypothetical protein Fisuc_2572 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76157.1
12	60	hypothetical protein Fisuc_2573 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76158.1
13		No significant similarities observed
14	46	protein of unknown function DUF323 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76159.1
15		No significant similarities observed
16	42	hypothetical protein Fisuc_2578 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76163.1
17	**41**	**Cellulase [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75598.1**
18	81	WD40 domain protein beta Propeller [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX73981.1
19	89	tRNA/rRNA methyltransferase (SpoU) [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX73977.1
20	83	hypothetical protein Fisuc_2766 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76349.1
21	37	cytochrome c assembly protein [*Planctomyces limnophilus* DSM 3776] >gb|ADG65876.1
22		No significant similarities observed
23		No significant similarities observed
24	61	PDZ/DHR/GLGF domain protein [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76110.1
25		No significant similarities observed
26	51	51% protein of unknown function DUF323 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76159.1
27		No significant similarities observed
28		No significant similarities observed
29		No significant similarities observed
30		No significant similarities observed
31	46	protein of unknown function DUF323 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76159.1
32		No significant similarities observed
33		No significant similarities observed
34	31	hypothetical protein Fisuc_1601 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75196.1
35		No significant similarities observed
36	77	Organic solvent tolerance protein OstA-like protein [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76111.1
37	42	protein of unknown function DUF323 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76161.1

3.5.2.2 Sequences of further fosmids derived from the library constructed with DNA from elephant faeces

Additional to pFosCelA84, the sequences of several fosmids were obtained by pyrosequencing (2.11.2). Many putative proteins showed highest identity to *Fibrobacter succinogenes* (Fig. 3.11 to 3.13). All fragments included at least one ORF coding for a putative protein involved in cell wall degradation or binding (Fig. 3.11 to 3.13).

Fosmid clone 82F10 was highly active in the presence of different ILs (Table 3.2). On this fosmid ORFs putatively coding for two carbohydrate binding family 6 proteins, an endoglucanase with highest identity to *Dictyoglomus thermophilum* and a cell wall binding repeat-containing protein with 26% identity to *Clostridium beijerinckii*, were detected (Fig. 3.12).

On fosmid 89E1 two ORFs putatively coding for glycoside hydrolases were discovered, a GH3 domain protein with highest identity to *Pedobacter heparinus* and a cellodextrinase with highest identity to an uncultured microorganism (Fig. 3.11). Furthermore a hypothetical protein which appeared to be a beta-glucanase was detected. This putative protein was closest related to *Bacteroides dorei*. The clone carrying fosmid 89E1 exhibited high activity against CMC but lost activity drastically with the addition of 30% (vol/vol) IL (Table 3.2, 3.3).

Fosmid clone 94A1 showed high activity in the assays without ILs but lost almost its complete activity in the presence of 30% (vol/vol) IL (Table 3.2, 3.3). Most ORFs detected on fosmid 94A1 putatively coded for proteins with highest identity to proteins from *F. succinogenes* (Fig. 3.13). On this fosmid ORFs were discovered that putatively coded for an endo-1,4-beta-glucanase and for a carbohydrate binding family 6 protein (Fig. 3.13).

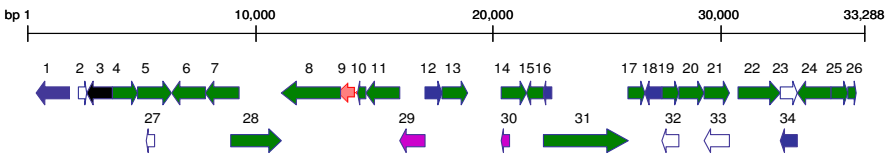

Figure 3.11: Putative ORFs discovered on fosmid 89E1

The phylum of the next known relative (aa identity) is indicated by the colour of the arrow. Green: *Bacteroidetes*, red: *Fibrobacteres*, blue: *Firmicutes*, purple: *Actinobacteria*, white: no significant similarities observed. All putative proteins are listed below in Table 3.5 with their closest relative (% aa identity).

Table 3.5: Putative proteins coded for on fosmid 89E1. Proteins involved in carbohydrate binding and hydrolysis are marked in bold.

ORF No.	aa identity (%) to the next known relative	next known relative
1	27	hypothetical protein CLOBOL_01602 [*Clostridium bolteae* ATCC BAA-613] >gb|EDP18247.1
2		No significant similarities observed
3	45	NADH pyrophosphatase [*Treponema vincentii* ATCC 35580] >gb|EEV20326.1
4	57	hypothetical protein BACINT_03215 [*Bacteroides intestinalis* DSM 17393] >gb|EDV04088.1
5	40	hypothetical protein BACUNI_03071 [*Bacteroides uniformis* ATCC 8492] >gb|EDO53056.1
6	41	conserved hypothetical protein [*Bacteroides* sp. 3_2_5] >ref|ZP_06094020.1
7	29	hypothetical protein PREME0002_2579 [*Prevotella melaninogenica* ATCC 25845] >gb|EES84168.1
8	**47**	**47% glycoside hydrolase family 3 domain protein [*Pedobacter heparinus* DSM 2366] >gb|ACU05829.1**
9	46	hypothetical protein Fisuc_0411 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX74023.1
10	61	transcription elongation factor GreA family protein [*Prevotella ruminicola* 23] >gb|ADE81651.1
11	**71**	**cellodextrinase [uncultured microorganism] >gb|ADB80112.1**
12	22	transcriptional regulator [*Mollicutes* bacterium D7] >gb|EEO32882.1
13	33	aminopeptidase [*Cytophaga hutchinsonii* ATCC 33406] >gb|ABG57326.1
14	60	hypothetical protein BACCOP_03778 [*Bacteroides coprocola* DSM 17136] >gb|EDU99212.1
15	45	hypothetical protein BACDOR_00314 [*Bacteroides dorei* DSM 17855] >gb|EEB27188.1
16	75	stress responsive A/B barrel domain protein *[Faecalibacterium* cf. *prausnitzii* KLE1255] >gb|EFQ06042.1
17	53	conserved hypothetical protein [*Prevotella* sp. oral taxon 472 str. F0295] >gb|EEX51538.1
18	37	hypothetical protein EUBSIR_01294 [*Eubacterium siraeum* DSM 15702] >gb|EDS00810.1
19	51	hypothetical protein ALIPUT_00913 [*Alistipes putredinis* DSM 17216] >gb|EDS03856.1
20	64	hypothetical protein ALIPUT_00914 [*Alistipes putredinis* DSM 17216] >gb|EDS03857.1
21	48	Pyridine nucleotide-disulphide oxidoreductase [*Alistipes shahii* WAL 8301] >emb|CBK64845.1
22	38	Fe-S oxidoreductase [*Alistipes shahii* WAL 8301] >emb|CBK64843.1
23		No significant similarities observed
24	59	hypothetical protein ALIPUT_01961 [*Alistipes putredinis* DSM 17216] >gb|EDS02435.1
25	40	hydrolase, haloacid dehalogenase-like family protein [*Flavobacteriaceae* bacterium 3519-10] >gb|ACU07340.1
26	42	hypothetical protein MED152_12649 [*Polaribacter* sp. MED152] >gb|EAQ40886.1
27		No significant similarities observed
28	37	hypothetical protein BACINT_01116 [*Bacteroides intestinalis* DSM 17393] >gb|EDV06031.1
29	57	hypothetical protein Shel_16430 [*Slackia heliotrinireducens* DSM 20476] >gb|ACV22662.1
30	22	ABC transporter related protein [*Streptomyces flavogriseus* ATCC 33331] >gb|EEW74504.1
31	**35**	**hypothetical protein BACDOR_03694 [*Bacteroides dorei* DSM 17855] >ref|ZP_04540678.1**
32		No significant similarities observed
33		No significant similarities observed
34	36	phosphohydrolase, MutT/nudix family protein [*Clostridium butyricum* E4 str. BoNT E BL5262] >gb|EEP56565.1

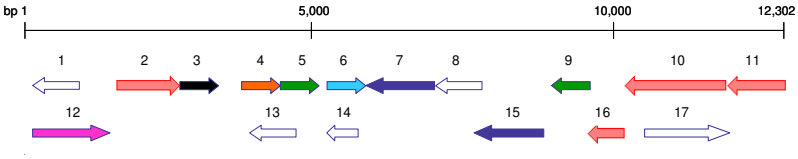

Figure 3.12: Putative ORFs discovered on fosmid 82F10

The phylum of the next known relative (aa identity) is indicated by the colour of the arrow. Green: *Bacteroidetes*, red: *Fibrobacteres*, black: *Spirochaetes*, orange: *Euryarchaeota*, blue: *Firmicutes*, pink: *Dictyoglomi*, white: no significant similarities observed. All putative proteins are listed below in Table 3.6 with their closest relative (% aa identity).

Table 3.6: Putative proteins coded for on fosmid 82F10. Proteins involved in carbohydrate binding and hydrolysis are marked in bold.

ORF No.	aa identity (%) to the next known relative	next known relative
1		No significant similarities observed
2	87	hypothetical protein Fisuc_2099 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75686.1
3	69	conserved hypothetical protein [*Treponema vincentii* ATCC 35580] >gb|EEV21130.1
4	60	hypothetical protein Mlab_0808 [*Methanocorpusculum labreanum* Z] >gb|ABN06979.1
5	64	protein of hypothetical function DUF763 [*Chryseobacterium gleum* ATCC 35910] >gb|EFK33774.1
6	49	protein of unknown function DUF763 [*Acidobacterium* sp. MP5ACTX8] >gb|EFI54033.1
7	35	hypothetical protein GK0744 [*Geobacillus kaustophilus* HTA426] >dbj|BAD75029.1
8		No significant similarities observed
9	25	hypothetical protein [*Alistipes shahii* WAL 8301] >emb|CBK62984.1
10	**60**	**Carbohydrate binding family 6 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76120.1**
11	**56**	**Carbohydrate binding family 6 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76120.1**
12	**44**	**endoglucanase H [*Dictyoglomus thermophilum* H-6-12] >gb|ACI19692**
13		No significant similarities observed
14		No significant similarities observed
15	**26**	**cell wall binding repeat-containing protein [*Clostridium beijerinckii* NCIMB 8052] >gb|ABR36498.1**
16	45	hypothetical protein Fisuc_1605 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75200.1
17		No significant similarities observed

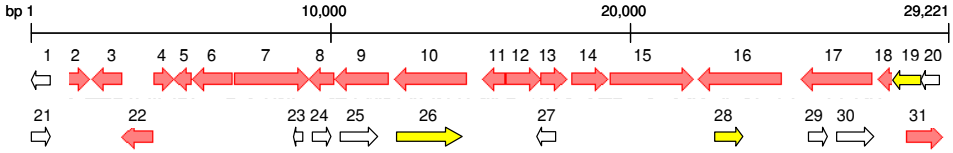

Figure 3.13: Putative ORFs discovered on fosmid 94A1

The phylum of the next known relative (aa identity) is indicated by the colour of the arrow. Yellow: *Proteobacteria*, red: *Fibrobacteres*, white: no significant similarities observed. All putative proteins are listed below in Table 3.7 with their closest relative (% aa identity).

Table 3.7: Putative proteins coded for on fosmid 94A1. Proteins involved in carbohydrate binding and hydrolysis are marked in bold.

ORF No.	aa identity (%) to the next known relative	next known relative
1		No significant similarities observed
2	81	hypothetical protein Fisuc_2765 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX76348.1
3	47	asparaginase [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX74200.1
4	37	hypothetical protein Fisuc_1835 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75427.1
5	88	GTP-binding protein [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75869.1
6	39	hypothetical protein Fisuc_2282 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75868.1
7	60	DNA internalization-related competence protein ComEC/Rec2 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75867.1
8	80	hypothetical protein Fisuc_2319 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75905.1
9	47	metal dependent phosphohydrolase [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75976.1
10	90	polyribonucleotide nucleotidyltransferase [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75974.1
11	47	hypothetical protein Fisuc_2278 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75864.1
12	73	histidinol-phosphate aminotransferase [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75865.1
13	79	histidinol-phosphate phosphatase [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75866.1
14	34	hypothetical protein Fisuc_2318 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75904.1
15	**65**	**endo-1,4-beta-glucanase/xyloglucanase, putative, gly74A [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75903.1**
16	88	ATPase AAA-2 domain protein [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75863.1
17	**69**	**Carbohydrate binding family 6 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75357.1**
18	41	hypothetical protein Fisuc_2114 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75701.1
19	31	hypothetical protein STIAU_4956 [*Stigmatella aurantiaca* DW4/3-1] >gb|EAU67275.1
20		No significant similarities observed
21		No significant similarities observed
22	42	hypothetical protein Fisuc_0589 [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX74201.1
23		No significant similarities observed
24		No significant similarities observed
25		No significant similarities observed
26	32	hypothetical protein EcE24377A_3648 [*Escherichia coli* E24377A] >gb|ABV19590.1
27		No significant similarities observed
28	26	hypothetical protein ISM_07185 [*Roseovarius nubinhibens* ISM] >gb|EAP78060.1
29		No significant similarities observed
30		No significant similarities observed
31	89	queuine tRNA-ribosyltransferase [*Fibrobacter succinogenes* subsp. *succinogenes* S85] >gb|ACX75951.1

3.6 Cloning of celA84

The cellulase gene encoded by pFosCelA84 was designated as *celA84* and was amplified (2.8) with the primer pair *celA84-Bam*HI-For and *celA84-Sal*I-Rev (2.1). The resulting fragment was ligated into pDrive (2.9.3.1) and transformed in *E. coli* DH5α (2.1). From the subsequently purified plasmid (2.5.3) the fragment comprising *celA84* was excised with *Bam*HI and *Sal*I (2.9.1). This fragment was then prepared from an agarose gel (2.5.5) and ligated (2.9.3.2) into the previously with *Bam*HI and *Sal*I linearised (2.9.1) and dephosphorylated (2.9.2) vector pMAL-c2 (2.1). The construct was transformed in *E. coli* BL21 (DE3) (2.1). Clones were analysed via direct colony PCR (2.8.4) and plasmid isolation (2.5.4) with subsequent restriction analysis (2.9.1).

3.7 Purification of CelA2, CelA3 and CelA84

In this study CelA2 (2.3.3), CelA3 (2.3.3) and CelA84 were purified. The gene *celA2* was ligated into pET28a (2.1), *celA3* and *celA84* in pMAL-c2 (2.1).

3.7.1 Purification of CelA2 with His-tag affinity column

The gene *celA2* was ligated into pET28a (2.1). Consequentially the produced protein contained an N-terminal His-tag sequence. The protein could therefore be purified with Ni-TED columns (2.13.4.1).

For the overexpression of *celA2*, an *E. coli* BL21 (DE3) pRIL (2.1) culture carrying pET28a-*celA2* was grown in LB medium (2.2.2) containing chloramphenicol and kanamycin (2.2.1) to an OD_{600} of 0.8 at 37°C. Induction of the enzyme's production was performed with the addition of 0.1 mM IPTG. The culture was then incubated for 3 h at 37°C.

When CelA2 was purified with His-tag affinity columns as described (2.13.4.1), several bands additional to the desired protein were detectable with SDS-PAGE (data not shown). Therefore 10 mM imidazole was added to the standard washing buffer. Then CelA2 could be purified to high homogeneity (Fig. 3.14). The molecular weight determined with SDS-PAGE of about 75 kDa was in accordance with the theoretical molecular weight of 69.03 kDa (Fig. 3.14).

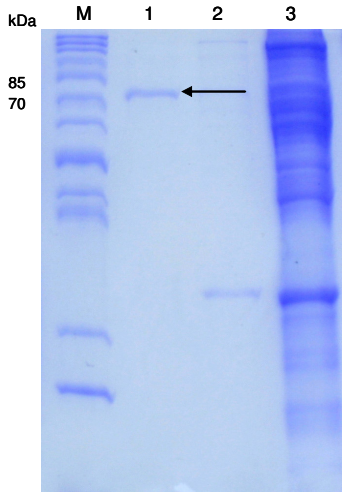

Figure 3.14: Purification of recombinant CelA2 from *E. coli* crude cell extracts

M: Page Ruler Protein ladder SM0661 (Fermentas, St. Leon-Rot, Germany), lane 1: 0.5 µg purified protein (CelA2 is shown with the arrow), lane 2: washing fraction, lane 3: crude extract

3.7.2 Purification with amylose-resin

As the genes *celA3* and *celA84* were ligated into pMAL-c2 (2.1), the respective proteins were produced as fusion proteins with an N-terminal maltose binding protein domain (MBP) and could therefore be purified with amylose-resin and eluted from the resin with maltose (2.13.4.2).

3.7.2.1 Purification of CelA3

To purify CelA3, an *E. coli* BL21 (DE3) pRIL (2.1) culture carrying pMAL-c2-*celA3* was grown in rich medium supplied with ampicillin and chloramphenicol (2.2.1, 2.2.3) to an OD_{600} of 0.5 at 37°C. The induction of the production of CelA3 was performed with the addition of 0.3 mM IPTG and the culture was furthermore incubated for 3 h at 37°C.

CelA3 could be purified as described above (2.13.4.2). The molecular weight ascertained by SDS-PAGE was about 125 kDa, which included the MBP (42.5 kDa) and CelA3 (theoretical molecular weight 69.95 kDa) (Fig. 3.15).

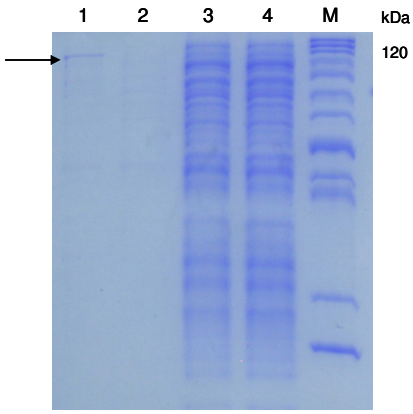

Figure 3.15: Purification of recombinant CelA3 from *E. coli* crude extracts

Lane 1: 1.1 µg purified protein (CelA3 is shown with the arrow), lane 2: washing fraction, lane 3: flow-through, lane 4 crude extract, M: Page Ruler Protein ladder SM0661 (Fermentas, St. Leon-Rot, Germany).

3.7.2.2 Purification of CelA84

For the production of CelA84, an *E. coli* BL21 (DE3) (2.1) culture with pMAL-c2-*celA84* was grown in rich medium supplied with ampicillin (2.2.1, 2.2.3) to an OD_{600} of 0.5 at 22°C. The gene's overexpression was induced with the addition of 0.3 mM IPTG and the culture was then incubated for 3 h at 22°C.

When CelA84 was purified as described above (2.13.4.2), several bands were visible on the polyacrylamide gel (data not shown). To avoid this problem, the MBP was cut from CelA84 while the fusion protein was bound to the amylose resin. Therefore the cell lysate was transferred onto the equilibrated column. After washing the column with column buffer, the column was allowed to empty until the resin was covered with a thin layer of buffer. Then 1 µL FXa (protease cleaving the MBP from the rest of the fusion protein, New England Biolabs (Beverly, MA, USA)) was added and the column was incubated at room temperature overnight. The next day proteins were eluted with column buffer. Via this method, CelA84 could be purified to high homogeneity (Fig. 3.16). The molecular weight determined via SDS-PAGE, 55 kDa (Fig. 3.16), was in accordance with the theoretical molecular weight of 54.37 kDa.

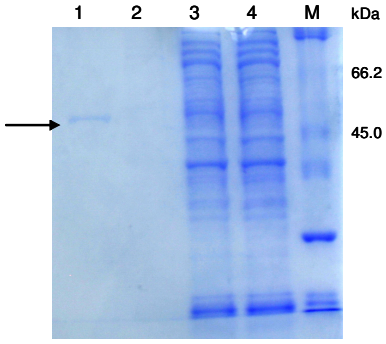

Figure 3.16: Purification of recombinant CelA84 from *E. coli* crude cell extracts

Lane 1: 1.7 µg purified protein (CelA84 is shown with the arrow), lane 2: washing fraction, lane 3: flow-through, lane 4 crude extract, M: Protein molecular weight marker SM0431 (Fermentas, St. Leon-Rot, Germany).

3.8 Characterisation of recombinant proteins

The characterisation of the recombinant cellulases was performed applying the DNSA assay (2.14.2). If not mentioned otherwise, CMC was used as a substrate. The cellulases were investigated for their pH- and temperature-optima as well as for their specific activity in the absence and presence of ILs. As the three described and purified IL-tolerant enzymes, derived from the metagenomes of the biogas plant (CelA2 and CelA3) and of the elephant faeces (CelA84), shared the feature of moderate to high thermostability, they were benchmarked with two further thermostable enzymes, CelA from *Thermotoga maritima* (Liebl et al. 1996) and Cel5K from a metagenomic library from the Avachinsky Crater in Kamchatka (2.3.4).

3.8.1 pH-optimum

The pH-optimum of CelA84 was determined with the DNSA assay (2.14.2). The standard buffer was replaced with McIlIvaine buffer with pH values of 5.5 and 6.5 or with 150 mM Tris pH 7.5 and 8.5. The pH-optimum of CelA84 was at pH 6.5 (Fig. 3.17).

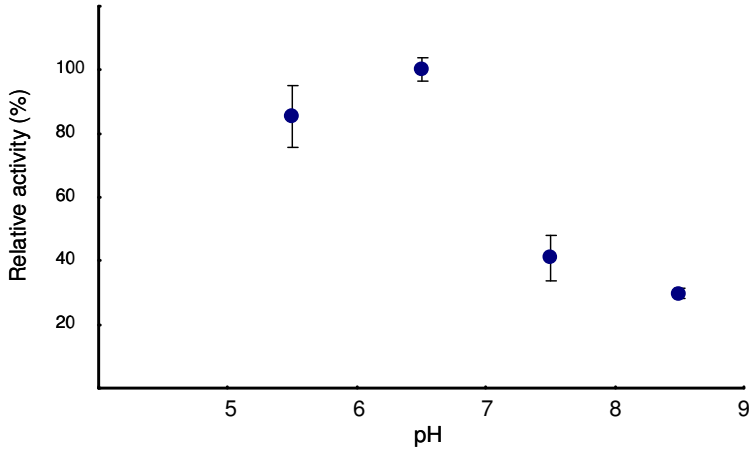

Figure 3.17: Relative activity of CelA84 at different pH values

McIllvaine buffer was used at pH 5.5 and 6.5, and at 7.5 and 8.5 150 mM Tris was applied.

3.8.2 Temperature optima

The temperature optima of CelA2, CelA3 and CelA84 were assayed in the standard buffer system and in the presence of 30% (vol/vol) [BMPL][OTF] and [EMIM][OTF] with the DNSA assay (2.14.2).

The temperature optimum of CelA2 without the addition of IL was 55°C (Fig. 3.18.A); this value was reduced significantly to 38°C in the presence of 30% (vol/vol) [BMPL][OTF] and [EMIM][OTF] (Fig. 3.18.A).

CelA3 was most active at 70°C (Fig. 3.18.B). In the presence of 30% (vol/vol) IL the temperature optimum was reduced to 66°C ([BMPL][OTF]) or 60°C ([EMIM][OTF]), respectively. In the presence of [EMIM][OTF] the activity curve was very even, the enzyme showed comparable activity over a wide temperature range (Fig. 3.18.B).

The temperature optimum of CelA84 without the addition of IL was 46°C (Fig. 3.18.C). The optimum temperature in the presence of both tested ILs was remarkably lower. In the presence of 30% (vol/vol) [BMPL][OTF] the temperature optimum was at 25°C and in the presence of 30% (vol/vol) [EMIM][OTF] it was at 31°C (Fig. 3.18.C). At the optimum temperature determined in buffer, 46°C, CelA84 was not active in the presence of any of the tested ILs (Fig. 3.18C, 3.19).

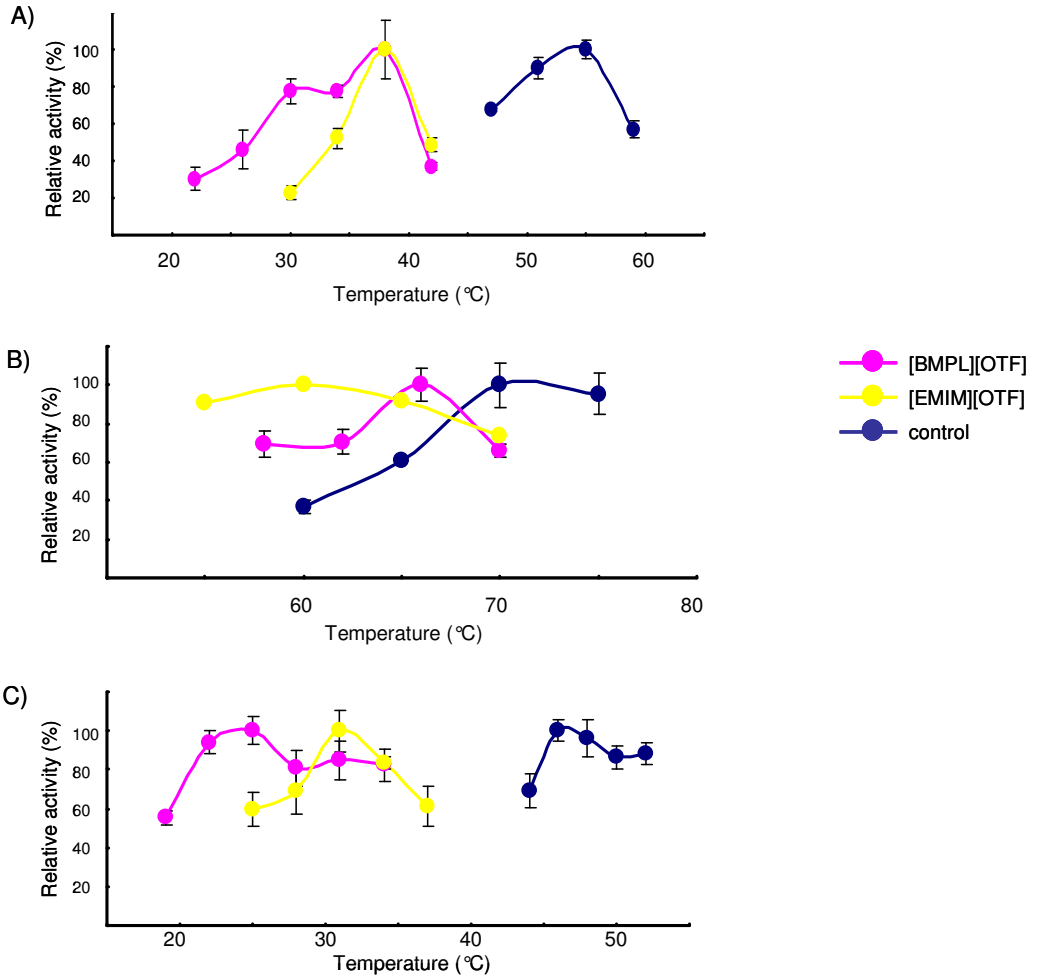

Figure 3.18: Determination of the temperature optima of A) CelA2, B) CelA3 and C) CelA84 in the presence of 30% (vol/vol) [BMPL][OTF], 30% (vol/vol) [EMIM][OTF] and without IL

3.8.3 Temperature stability

The thermostability was tested for CelA2, CelA3, CelA84 and CelA at the respective temperature optimum; Cel5K was assayed at 70°C. Therefore the standard assay without substrate was incubated at the respective temperature. After a specific time CMC was added and the activity was measured as described for the standard assay (2.14.2). Over a time period of 4 to 5 days, CelA2 and CelA did not lose activity remarkably at the respective temperature optima of 55°C and 83°C (data not shown). CelA84 lost 43% activity after one day at 46°C and Cel5K exhibited 65% activity after 3 days at 70°C. CelA3 showed 13% relative activity after 5 days at 70°C (data not shown).

3.8.4 Specific activity against CMC

The enzymes' specific activities were evaluated at 37°C and at the respective optimum temperature (Cel5K was assayed at 70°C instead of 96°C) with the standard DNSA assay (2.14.2).

At both temperatures CelA2 was the most active enzyme, its specific activity was 6.66 U/mg at 37°C and 11.94 U/mg at 55°C while CelA3 was the least active enzyme (0.64 U/mg at 37°C and 1.87 U/mg at 70°C) (Fig. 3.19). The specific activity of CelA84 was 5.16 U/mg at 37°C and 7.48 U/mg at 46°C. Cel5K showed the highest difference between the activities at both temperatures (8.05 U/mg at 70°C and 1.13 U/mg at 37°C) while the specific activity of CelA was 4.82 U/mg at 37°C and 8.49 U/mg at 83°C (Fig. 3.19).

3.8.5 Activity in the presence of 30% (vol/vol) IL

The activity of all five cellulases was furthermore assayed in the presence of 30% (vol/vol) of the different ILs at 37°C and at the respective optimum temperature (Cel5K was assayed at 70°C instead of 96°C).

In general, all cellulases were highly active in the presence of 30% (vol/vol) of several ILs (Fig. 3.19). Altogether, the highest activities in the presence of ILs were observed for CelA at 83°C. At this temperature, CelA was even more active in the presence of 30% (vol/vol) [EMIM][ATF] (9.40 U/mg) and [EMIM][OTF] (11.15 U/mg) than in the control assay (8.49 U/mg). At 37°C as well as at 70°C Cel5K exhibited about the same activity in the presence of [BMIM][OTF], [EMIM][ATF], [EMIM][OTF] and in the control (6.5 to 8 U/mg at 70°C and 1 to 1.6 U/mg at 37°C) (Fig. 3.19). In contrast, CelA3 showed comparable activity in the presence of all ILs and in the control (1 to 2 U/mg at 70°C and 0.4 to 0.8 U/mg at 37°C) (Fig. 3.19). CelA2 also exhibited high activity in the presence of 30% (vol/vol) of some of the investigated ILs, the highest activities were measured in the presence of the ILs [BMIM][OTF] (3.86 U/mg), [BMPL][OTF] (4.86 U/mg) and [EMIM][OTF] (6.37 U/mg) at 55°C (Fig. 3.19). Except CelA84, all cellulases were more active in the presence of ILs at the respective optimum temperature than at 37°C. In contrast, CelA84 lost almost its complete activity at 46°C in the presence of all investigated ILs while it showed about 0.5 U/mg at 37°C in the presence of 30% (vol/vol) [BMIM][OTF], [EMIM][ATF], [BMPL][OTF] and [EMIM][OTF] (Fig. 3.19). An additional

observation was that CelA2, CelA and CelA84 shared the feature of being almost inactive in the presence of the chloride containing ILs, [BMIM]Cl and [BMMIM]Cl (Fig. 3.19)

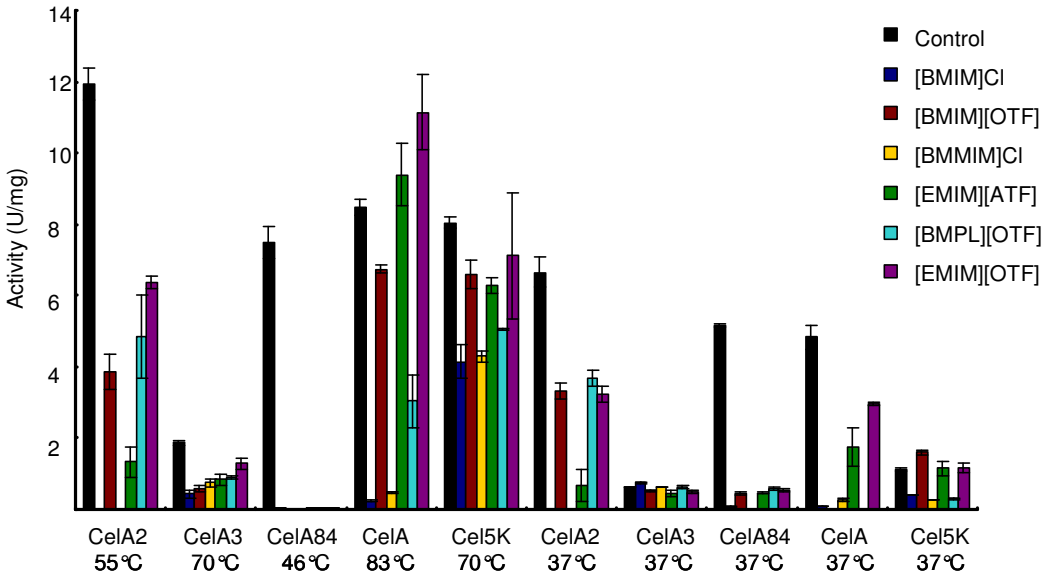

Figure 3.19: Activity of the investigated recombinant and purified enzymes in the presence of 30% (vol/vol) of the different ILs at 37℃ and the respective temperature optimum (Cel5K was assayed at 70℃ instead of 96℃)

3.8.6 Long term stability in the presence of ILs

The long term stability of the enzymes was tested in the presence of 60% (vol/vol) IL. Thereby the IL concentration in the activity measurement was 30% (vol/vol) after the addition of the substrate CMC. The assays were incubated for several days at room temperature or at the respective optimum temperature, while the residual activity was assayed at 37℃ (after incubation at room temperature) or the respective optimum temperature, respectively (Cel5K was assayed at 70℃ instead of 96℃).

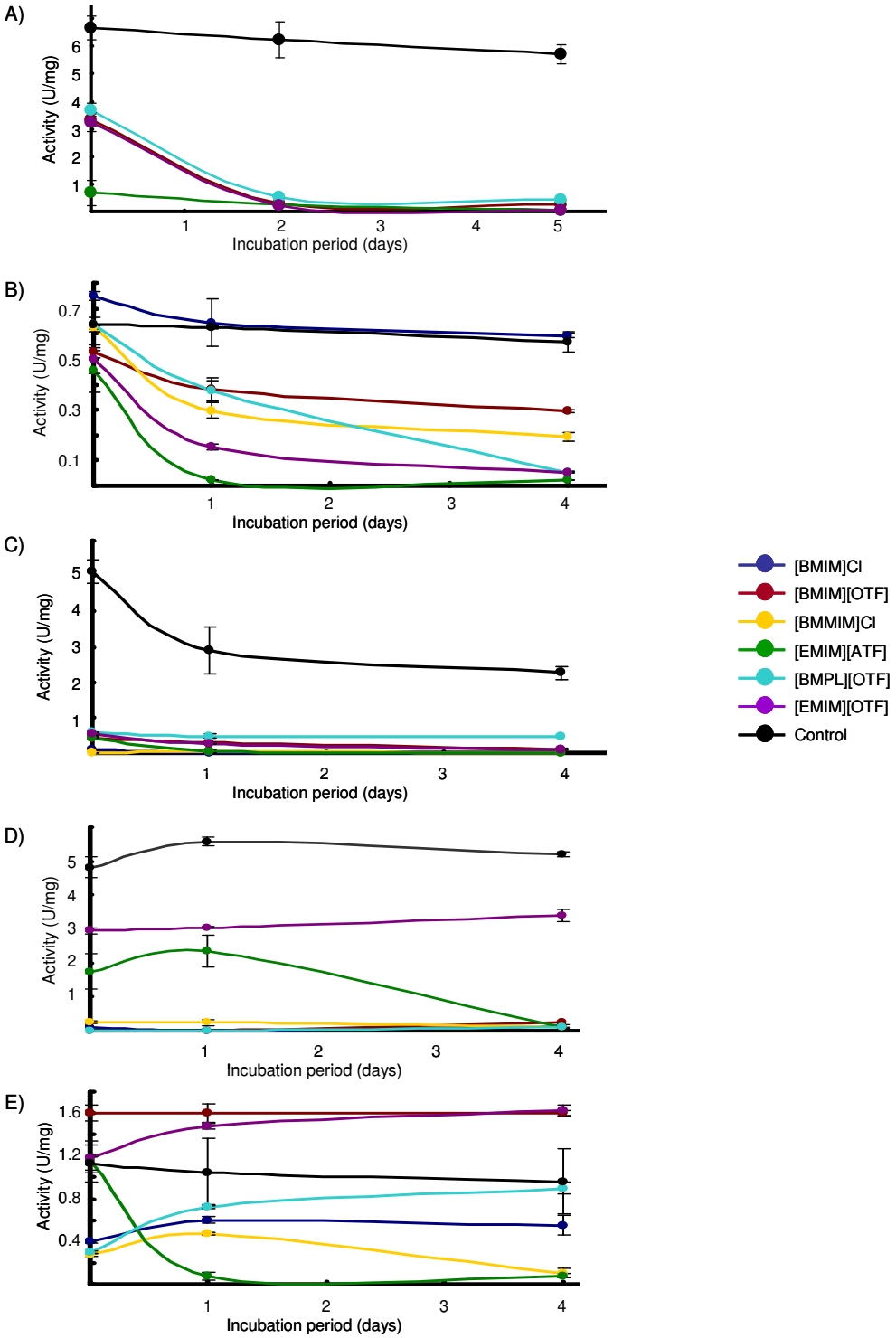

Figure 3.20: Activity of A) CelA2, B) CelA3, C) CelA84, D) CelA and E) Cel5K in the presence of 30% (vol/vol) IL at 37°C after pre-incubation in 60% (vol/vol) IL at room temperature

When the assays were incubated at room temperature, all cellulases, except CelA2, stayed active in the presence of at least one of the tested ILs after 4 days (Fig. 3.20). The activity of CelA2 was reduced to maximal 10% of the initial activity in the presence of the respective IL after an incubation period of 5 days ([BMIM][OTF] and [BMPL][OTF], Fig. 3.20.A). CelA84 and CelA were stable in the presence of one IL each; CelA84 retained 81% of the initial activity in the presence of [BMPL][OTF] (Fig. 3.20.C) while CelA showed 115% relative activity after 4 days in the presence of [EMIM][OTF] (Fig. 3.20.D). CelA3 and Cel5K stayed active in the presence of several ILs after 4 days at room temperature. Cel5K was stable in the presence of [BMIM]Cl, [BMIM][OTF], [BMPL][OTF] and [EMIM][OTF]. In the presence of all these ILs, the activity after 4 days was at least 100% of the initial activity in the presence of the respective IL (Fig. 3.20.E). CelA3 still exhibited remarkable relative activity after 4 days in the presence of [BMIM]Cl (79%), [BMIM][OTF] (56%) and [BMMIM]Cl (31%) (Fig. 3.20.B). None of the cellulases was stable in the presence of [EMIM][ATF], though the activity of all cellulases in the presence of this IL was relatively high (Fig. 3.20).

After several days at the optimum temperatures (Cel5K was investigated at 70°C instead of 96°C), the activity of all cellulases was reduced significantly in the presence of all ILs. None of the cellulases exhibited remarkable activity in the presence of any of the ILs anymore (Fig. 3.21). CelA84 did not even show activity in the presence of any of the tested ILs at 46°C without pre-incubation (Fig. 3.21.C).

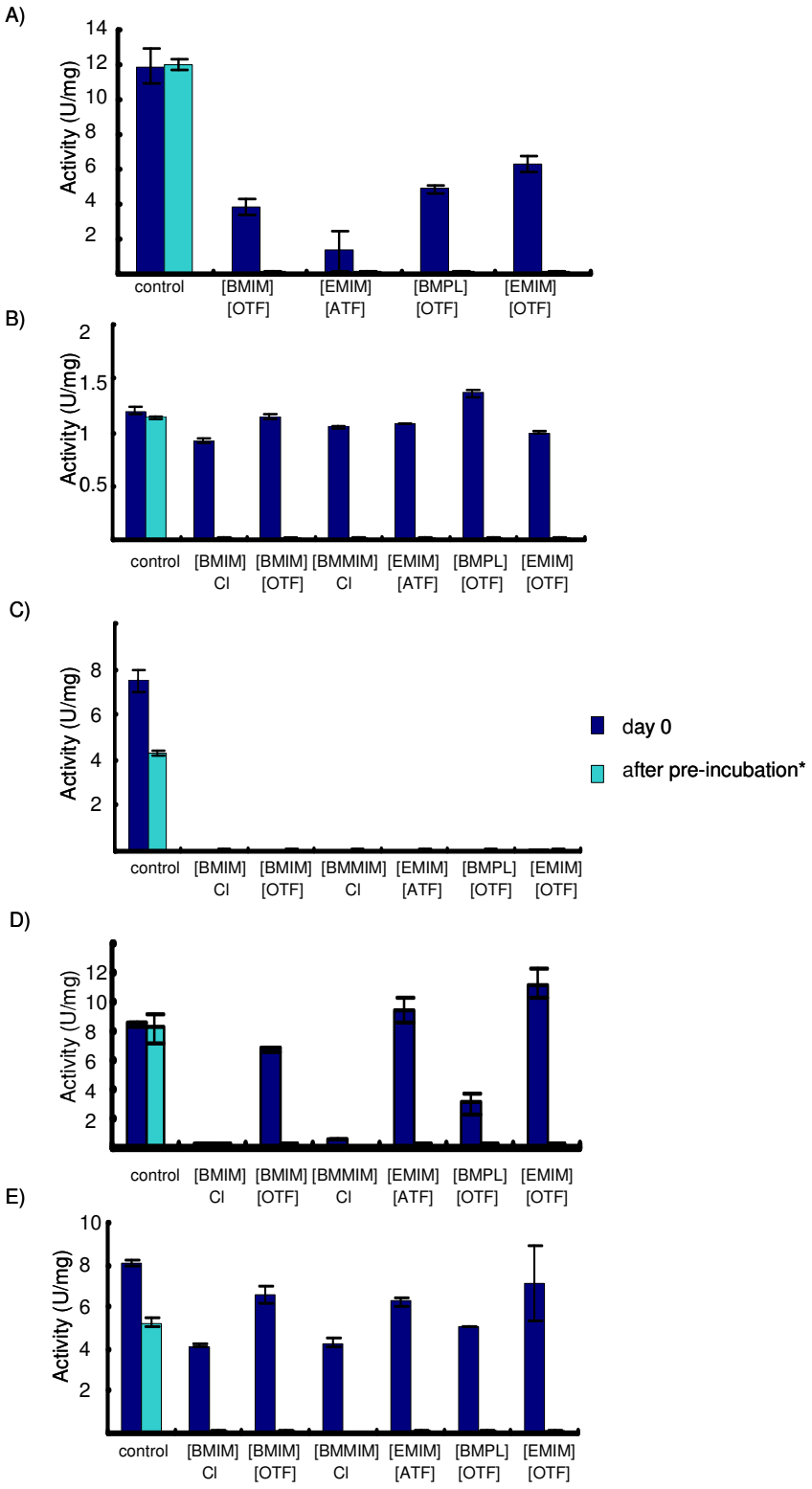

Fig. 3.21: Activity of A) CelA2, B) CelA3, C) CelA84, D) CelA and E) Cel5K in the presence of 30% IL after pre-incubation in 60% IL at A) 55°C, B) 55°C, C) 46°C, D) 83°C and E) 70°C for *A) 2 days, B) 3 days, C) 1 day, D) 3 days and E) 3 days

3.8.7 Substrate specifics of CelA2, CelA3, CelA84, CelA and Cel5K

The cellulases' activity was additionally tested against crystalline cellulose and crystalline cellulose that had been dissolved in 1-butyl-3-methylimidazolium acetate ([BMIM]Ac) or 1-ethyl-3-methylimidazolium acetate ([EMIM]Ac) and afterwards precipitated with the addition of water (2.14.2.4). In these assays the IL concentration was 4.32% (w/vol).

None of the cellulases was active against crystalline cellulose (data not shown), but CelA and Cel5K were able to hydrolyse the cellulose that had been dissolved in IL and precipitated with water. Both enzymes showed higher activity against cellulose dissolved in [BMIM]Ac than in [EMIM]Ac (Fig. 3.22). While Cel5K was able to hydrolyse this substrate at 37°C as well as at 70°C, CelA was only active against this substrate at 37°C (Fig. 3.22).

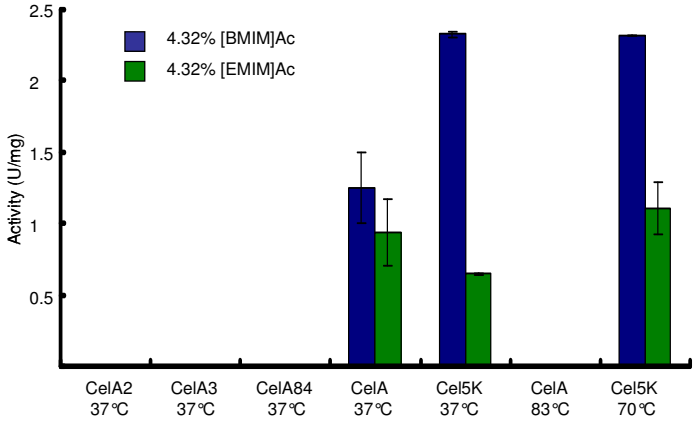

Figure 3.22: Activity of the investigated enzymes against crystalline cellulose that was dissolved in [BMIM]Ac or [EMIM]Ac and then precipitated with water

3.9 Cloning of two further genes coding for GH5 cellulases

As two of the IL-tolerant cellulases showed highest identity to enzymes from *Clostridiales*, and three of the cellulases belonged to GH family 5, two further GH5 cellulases belonging to bacteria from *Clostridiales* were investigated. These enzymes were EngB from *Clostridium cellulovorans* (Acc. No. YP_003842670.1) and Cel5D from *Ruminococcus albus* (Acc. No. YP_004104242.1). *C. cellulovorans* and *R. albus* were grown anaerobically in the respective media (2.2.5, 2.2.6). Genomic DNA was isolated (2.5.1) and the

genes were amplified (2.8) with the primers *engB*-For and *engB*-Rev or *cel5D*-For and *cel5D*-Rev, respectively (2.1). The PCR products were ligated into pDrive (2.9.3.1) and the ligation was then transformed (2.10.2) in *E. coli* DH5α (2.1). The clones were verified with direct colony PCR (2.8.4) and plasmid purification (2.5.4) with a subsequent analytical digestion (2.9.1).

For the evaluation of the enzymes' activities in the presence of ILs (2.14.2.1) crude extract (2.13.1) of the enzyme clones, designated as pEngB and pCel5D, was employed. Both clones showed high activity against CMC, but in the presence of 30% (vol/vol) of the different ILs, none of them was active (Table 3.8).

Table 3.8: Activity of crude extracts of the clones pEngB and pCel5D in the presence of 30% (vol/vol) of different ILs

Clone	**GH**	**Activity (mU/mg)**						
		control	**[BMIM] Cl**	**[BMIM] [OTF]**	**[BMMIM] Cl**	**[EMIM] [ATF]**	**[BMPL] [OTF]**	**[EMIM] [OTF]**
pEngB	5	660±21	0	0	0	0.42±0	0	0
pCel5D	5	3260±91	0	0	0	0.06±0	0	0.30±0

4 Discussion

Cellulose is a valuable biopolymer that can be used as a source for glucose and to produce subsequent molecules like ethanol. The enzymatic process of cellulose hydrolysis has as major drawback the insolubility of crystalline cellulose in water. This is in conflict with the corresponding enzymes that need an aqueous medium for efficient hydrolysis. Alternative solvents are ionic liquids (ILs) as some of these are able to dissolve cellulose (Heinze et al. 2005, Swatloski et al. 2002). ILs could therefore be used as a reaction medium if cellulases were active and stable in their presence.

A tool for the discovery of novel enzymes is metagenomics (Schmeisser et al. 2007). This is a culture-independent technology allowing the investigation and utilisation of the complete diversity of microorganisms instead of only the cultured fraction of about 0.1 to 1% (Amann et al. 1995).

The aim of this study was to identify novel cellulases that are active and stable in the presence of high IL concentrations. Therefore metagenomic libraries were constructed and screened for cellulolytic clones. These clones were evaluated with regard to their activity and stability in the presence of different ILs. The most active enzymes were purified and characterised. As these enzymes were moderately thermostable, two further thermostable cellulases were investigated, CelA from *Thermotoga maritima* and a metagenomic cellulase derived from the Avachinsky Crater in Kamchatka designated as Cel5K.

4.1 Metagenomics as a tool for the discovery of industrially relevant cellulases

Metagenomics, i.e. the construction and screening of metagenomic libraries, has turned out to be a powerful tool for the discovery of novel enzymes and secondary metabolites like antibiotics as it enables the investigation of the complete microbial diversity and not only the cultivated proportion of about 0.1 to 1% (Amann et al. 1995). In several metagenomic studies cellulases with interesting properties have been detected. In a survey investigating metagenomic cellulases from soil, a glucanase active against a wide variety of

carbohydrates was detected (Liu et al. 2011). The metagenomic investigation of soil and water samples resulted in the discovery of several cellulases whereof one was moderately active but instable after an overnight incubation in the presence of ILs (Pottkämper et al. 2009).

Nevertheless, this technology also has its challenges. The choice of habitat is the first critical step. It has been shown that it is possible to discover novel enzymes from habitats that are phylogenetically diverse but not conditionally extreme or self-evident for the desired enzyme, e.g. cellulases were detected in soil and water samples (Kim et al. 2008, Pottkämper et al. 2009, Rondon et al. 2000). In contrast, many studies have been successful in the identification of novel enzymes by analysing extreme habitats like hot springs, where a thermostable esterase was discovered (Rhee et al. 2005) or obvious habitats like buffalo rumen (Duan et al. 2009) and compost (Pang et al. 2009) for the discovery of cellulases. Next to the investigation of self-evident habitats, enrichment cultures are a strategy to accumulate and investigate bacteria with the favoured activity. Furthermore the discovery of enzymes with the desired properties is enabled as pH, temperature and other conditions can be adjusted in the culture. The major drawback of this method is the loss of diversity as fast growing bacteria adapted to the respective culture conditions outcompete other bacteria (Daniel 2004). Nevertheless, different groups have been successful with this method, e. g. several dehydratases were identified in enrichments inoculated with water from the river Grone or a solar lake or with soil from a sugar beet field (Knietsch et al. 2003a).

Another critical step is the DNA purification. The chosen purification method biases the proportion of DNA that is extracted and therefore that of cloned and discovered genes. Most of the DNA extraction methods do not lyse all cells and therefore exclude parts of the bacterial DNA (Herrera and Cockell 2007, V. Wintzingerode et al. 1997). The choice of habitat and the DNA purification method furthermore influence the quality of the purified DNA. Different compounds which are co-purified with DNA, like humic acids from soil, hamper enzymatic DNA processing like digestion and ligation (Schmeisser et al. 2007).

A major bottleneck is the detection of the requested genes in metagenomic libraries. This can be performed using either function- or sequence-based screening methods. Sequence-based approaches like hybridisation methods or

PCR have the disadvantage that the probe and primer design depends on known sequences. Completely novel genes therefore cannot be detected. The major drawback of function-based approaches is the recovery of activity in the chosen host bacterium. It has been predicted that only about 40% of all enzymes can be recovered by their function when *E. coli* is used as a host (Gabor et al. 2004a). Reasons therefore can be found at different levels, e.g. transcription, translation and secretion. In detail problems can be different codon usages and varying sigma factors, transcription factors, ribosome binding sites and promoters. Missing cofactors, solubility and toxicity of recombinant proteins can be further difficulties (Baneyx 1999, Gabor et al. 2004a, Makrides 1996). For the correct folding, specific chaperones might be required. In some screening systems, proteins also need to be secreted to be detected by their activity. Secretion in foreign hosts often fails due to missing secretion systems or differing recognition sites (Baneyx 1999, Gabor et al. 2004a, Makrides 1996). Recently, several alternative host bacteria belonging to a wide range of phyla have been developed successfully. These are e.g. *Rhizobium leguminosarum* (Wexler et al. 2005), *Pseudomonas putida* (Martinez et al. 2004), *Desulfovibrio desulfuricans* and *D. fructosovorans* (Rousset et al. 1998), *Streptomyces lividans* (Martinez et al. 2004) and *Bacillus subtilis* (Murashima et al. 2002).
If these difficulties are circumvented, the identification, sequencing and cloning of the respective genes and the purification and characterisation of the enzymes are subsequent steps.

In this study, several habitats thought to comprise a broad diversity of cellulolytic microorganisms and consequentially enzymes were evaluated with regard to activity and biodiversity level prior to the construction of metagenomic libraries. The habitats chosen for this evaluation were elephant faeces, a biogas plant sample and enrichment cultures inoculated with elephant faeces, koala faeces, the biogas plant sample and with an extract of the shipworm *Teredo navalis*. In this study, all enrichments showed much lower activity against carboxymethylcellulose (CMC) than the habitats directly investigated, though activity was also measurable in the enrichments. The biogas plant sample, the

faeces of the Asian elephant and the *Teredo* enrichment were the most active habitats and were therefore chosen for further investigation.

The consortium of the enrichment culture inoculated with an extract of *Teredo navalis* showed an activity of about 2 mU/mg and was therefore the most active among the enrichment cultures. A 16S rRNA gene analysis revealed that this consortium mainly consisted of *γ-Proteobacteria*, and additionally of *α-* and *β-Proteobacteria*, *Firmicutes*, *Actinobacteria* and *Bacteroidetes*. *Teredo navalis* lives in sea water and exclusively feeds on wood. It possesses its own cellulases and is additionally inhabited by symbiotic cellulolytic bacteria (Distel et al. 1991). The enrichment culture inoculated with an extract of *Teredo navalis* therefore could have been highly hydrolytic, but it showed only poor activity. Probably the culture conditions or the substrate were not appropriate for the enrichment of bacteria active against CMC.

Faeces and intestinal samples are generally described as being very diverse (Dowd et al. 2008, Eckburg et al. 2005). The 16S rRNA gene analysis of the microorganisms in the faeces of the Asian elephant also exhibited very high biodiversity. The faeces comprised bacteria of all major phyla. Many sequences showed highest identity to the 16S rRNA gene of so far uncultivated bacteria. The investigation of this habitat hence allowed the discovery of novel enzymes. Elephants consume high amounts of cellulosic plant material like wood and grass and should therefore be inhabited by a high diversity of cellulolytic organisms. Consequentially, the cellulolytic activity of the metaproteome of the faeces sample was high, but the addition of 30% (vol/vol) ILs resulted in a loss of activity to about 10% relative activity.

The biogas plant sample derived from a reactor fed with cellulosic plant material, corn silage, run at a temperature of about 55°C. Under these conditions an anaerobic consortium that efficiently hydrolyses cellulose and that is furthermore thermostable can emerge. The microbiota of the biogas plant sample was rather homogeneous, it mainly comprised *Firmicutes* and some representatives of *Bacteroidetes* (Meske 2009). Nevertheless, the number of cellulases should be high in a biogas plant fuelled with corn silage. The cellulolytic activity of the metaproteome of this sample was also high and in the presence of 30% (vol/vol) IL it retained about 50% relative activity. Altogether,

the biogas plant sample and the elephant faeces sample seemed to be appropriate habitats for the discovery of a wide variety of cellulases.

The metagenomic libraries investigated in this study were fosmid libraries. The choice of a large insert vector is reasonable for the detection of cellulase genes, as these are often long (Warner et al. 2010) and many genes coding for cellulases of anaerobic bacteria are organised in gene clusters with additional genes coding for other cellulosomal proteins (Doi et al. 2003, Tamaru et al. 2000). Cellulosomes are large protein complexes consisting of a scaffold protein which is linked to the bacterial cell wall, to the substrate via a cellulose binding module (CBM) and to hydrolytic enzymes via specific cohesin-dockerin interactions (Fontes and Gilbert 2010). Putative cellulosomal elements have not been discovered in the sequences of the fosmids from the elephant faeces library that were obtained by pyrosequencing. The sequences of most of the putative proteins showed highest identity to proteins from *Fibrobacter succinogenes*. This bacterium does not produce cellulosomes but is known to contain genes coding for non-catalytic CBM proteins. The function of these proteins is unclear (Mitsumori et al. 1996, Suen et al. 2011). ORFs putatively coding for this kind of protein could be found in the sequences of some of the fosmids derived from the library constructed with DNA from elephant faeces. Additionally a variety of glycoside hydrolases, including cellodextrinases, endoglucanases and xyloglucanases, could be discovered displaying a highly hydrolytic community.

In the three investigated metagenomic libraries, derived from elephant faeces, the biogas plant sample and the *Teredo* enrichment, 14 cellulolytic clones were detected by their function using CMC as a substrate. Most of the clones showed high cellulolytic activity. Only the one clone from the library derived from the *Teredo* enrichment showed very weak activity corresponding to the weak activity of the respective enrichment culture. When evaluated for their activity in the presence of 30% (vol/vol) IL and after overnight incubation in 60% (vol/vol) IL, some of the cellulases lost almost their complete activity. Nevertheless, 79% of the investigated clones showed activity in the presence of 30% (vol/vol), and 57% remained active after overnight incubation in the

presence of 60% (vol/vol) of at least one of the ILs used in this study. In comparison with a previous study investigating mesophilic metagenomic cellulases from water and soil samples, where most of the enzyme clones lost almost their complete activity in the presence of 30% (vol/vol) IL (Pottkämper et al. 2009), a high proportion of IL-tolerant enzymes was detected. In conclusion, the chosen habitats, the biogas plant sample and elephant faeces, were appropriate for the discovery of IL-tolerant cellulases.

4.2 Characterisation of IL-tolerant cellulases

Two clones from the library derived from the biogas plant sample, designated as pFosCelA2 and pFosCelA3, and pFosCelA84 from the library constructed with DNA from elephant faeces were highly active and stable overnight in the presence of different ILs and were therefore chosen for further investigation. As the respective metagenome-derived enzymes, CelA2, CelA3 and CelA84, were IL-tolerant and moderately thermostable, they were benchmarked with two further thermostable cellulases; these were CelA from *Thermotoga maritima* (Liebl et al. 1996) and Cel5K from a metagenomic library derived from the Avachinsky Crater in Kamchatka.

4.2.1 Phylogeny and modular structure

The two enzymes derived from the library constructed with DNA from the biogas plant, CelA2 and CelA3, showed highest identity to enzymes from *Clostridia* (Table 4.1). CelA2 was closest related to EngO from *Clostridium cellulovorans* and CelA3 to an enzyme belonging to GH family 5 from *Clostridium cellulolyticum*. Many *Clostridia* are able to hydrolyse crystalline cellulose. They often possess cellulosomes and furthermore contain free cellulases (Doi et al. 1998, Doi and Tamaru 2001). The cellulases discovered in this study did not contain any dockerin domain that would bind to the cohesion of the respective scaffold protein. Hence it can be assumed that they are free enzymes.

CelA84, the enzyme derived from the library constructed with DNA from elephant faeces, was most identical to a cellulase from *Fibrobacter*

succinogenes. This is a cellulolytic bacterium often detected in intestinal tracts of herbivores that digests crystalline cellulose (Kobayashi et al. 2008).

On sequence level CelA2, CelA3 and CelA84 were all only weakly related to previously described enzymes. CelA2 and CelA84 shared 41% amino acid (aa) sequence identity to their respective next known relatives, CelA3 60% (Table 4.1). The low sequence identities indicated the novelty of the discovered enzymes. These data proved that metagenomics is an appropriate tool for the discovery of novel enzymes; this is additionally corroborated by the sequence data obtained via pyrosequencing, which exhibit many putative proteins with low sequence identities to their next known relatives. Furthermore the adequacy of function-based screening approaches for the identification of novel enzymes is demonstrated.

The IL-tolerant cellulases investigated in this study showed highest identity to enzymes from organisms that belong to completely different phylogenetic groups (Table 4.1). CelA2 and CelA3 were most closely related to enzymes from bacteria belonging to the class *Clostridia*, the others to bacteria from the phyla *Fibrobacteres* (CelA84), *Thermotogae* (CelA) and *Deinococcus-Thermus* (Cel5K). As two of the enzymes showed highest identity to enzymes from *Clostridia* as well as the previously described IL-tolerant cellulase from *Thermoanaerobacter tengcongensis* (Liang et al. 2011), two further cellulase genes belonging to *Clostridia*, *engB* from *Clostridium cellulovorans* and *cel5D* from *Ruminococcus albus*, were cloned. The corresponding enzymes did not show any activity in the presence of 30% (vol/vol) IL though their activity in the absence of ILs was high. Previously described moderately IL-tolerant cellulases belong to different phylogenetic groups as well, e.g. *γ-Proteobacteria* or *Euryarchaeota* (Datta et al. 2010, Pottkämper et al. 2009). Therefore the phylogenetic derivation of enzymes does not seem to correlate with their IL-tolerance.

The investigated enzymes belonged to GH family 5 (CelA3, CelA84 and Cel5K), 9 (CelA2) or 12 (CelA). It is noticeable that three of the enzymes belonged to GH family 5. This is a large group with 3,511 GenBank entries

(http://www.cazy.org/GH5.html) comprising enzymes from *Archaea*, *Bacteria* and *Eukarya.* These enzymes contain a $(\beta/\alpha)_8$ fold, also known as the TIM-barrel. This is a very common fold that has been described for a wide variety of enzymatic functions (Wierenga 2001). Glycoside hydrolases with this fold, as well as those with a β-jelly roll (e.g. enzymes from GH family 12), catalyse the hydrolysis of glycosidic bonds via the retaining mechanism (Yennamalli et al. 2011). In contrast the hydrolysis by enzymes with an $(\alpha/\alpha)_6$ fold, e.g. enzymes from GH family 9, proceeds via the inverting mechanism. This mechanism is not completely elucidated, but, in contrast to the retaining mechanism, the configuration of the anomeric carbon is inverted (Yennamalli et al. 2011). As the most IL-tolerant enzymes described previously, Cel5A, *Tma* Cel5A, *Pho* EG and CelA10, belong to GH5 as well (Datta et al. 2010, Liang et al. 2011, Pottkämper et al. 2009), it would be likely to speculate that enzymes of this family are predestined for being IL-tolerant. But, in contrast, the two furthermore investigated cellulases belonging to GH5, EngB and Cel5D, were inactive in the presence of 30% (vol/vol) IL. The reason for the detection of several IL-tolerant GH5 cellulases might thus be the high abundance of enzymes of this family.

Generally it seems unlikely that additional domains play a pivotal role in the determination of cellulases' activity and stability in the presence of ILs. CelA84, CelA and Cel5K did not comprise any additional domain. CelA2 contained an Ig-like domain. Ig-like domains have been described for several GH9 enzymes (Kataeva et al. 1999, Zverlov et al. 1998). The function of this domain is not completely elucidated, but interactions between the Ig-like and the catalytic domain have been described (Kataeva et al. 2004, Pereira et al. 2009). These interactions affect the fold and the stability of the enzyme, and the deletion of the Ig-like domain resulted in enzyme inactivation (Kataeva et al. 2004). CelA3 comprised two CBMs of the family 17_28. CBMs are domains that bind to carbohydrates, e.g. crystalline and amorphous cellulose, xylan, mannan and chitin and that are grouped in 39 families (Boraston et al. 2004). CBMs of the family 17_28 bind to amorphous cellulose (Araki et al. 2009). Previously described moderately IL-tolerant bacterial cellulases contained either a CBM family 2, a CBM family 17_28, a CBM family 10 and an Ig-like domain or no additional domain (Datta et al. 2010, Liang et al. 2011, Pottkämper et al. 2009).

Altogether it does not seem as if there was a correlation between the modular structure of cellulases and their tolerance against ILs. It has furthermore been described that CBMs do not affect the thermostability of enzymes (Anbarasan et al. 2010), what might be a feature linked to IL-tolerance.

Table 4.1: Characteristics of the IL-tolerant cellulases investigated in this study

Cellulase	GH family	Additional domains	No. of aa/ MW (kDa)	aa similarity/identity (%) to	Specific activity at $T_{opt.}$ (U/mg)
CelA2	9	Ig-like domain	604/69.03	56/41, EngO, *Clostridium cellulovorans*	11.9
CelA3	5	2 CBMs (family 17_28)	634/69.95	75/60, a GH5, *Clostridium cellulolyticum*	1.9
CelA84	5	-	506/54.37	55/41, cellulase, *Fibrobacter succinogenes*	5.1
CelA	12	-	257/29.73	*Thermotoga maritima*	8.5
Cel5K	5	-	346/38.67	74/61, cellulase, *Thermus caldophilus*	78.9

4.2.2 Halotolerance

Halotolerance is a feature that would logically be linked to tolerance against ILs in aqueous solution, as ILs dissociate into the respective ions in aqueous solution as inorganic salts do. The three metagenome-derived cellulases discovered in this study shared the feature of high tolerance against NaCl. CelA2 was the most stable enzyme, after 34 days in 4 M NaCl it still showed 80% relative activity (Table 4.2). CelA3 and CelA84 were somewhat less stable; they retained about 50% relative activity after the same time period in 4 M NaCl (Table 4.2). CelA from *T. maritima*, which has also been shown to be IL-tolerant in this study, was even more active in NaCl concentrations of 5 M at elevated temperatures (85°C and 95°C) than at lower NaCl concentrations (Liebl et al. 1996). All enzymes that were shown to be at least moderately IL-tolerant and that have been tested for their halotolerance showed high activity in the presence of NaCl (Table 4.2), e.g. the cellulase from *Thermoanaerobacter tengcongensis* MB4 showed 50% activity in 3 M NaCl (Liang et al. 2011). The

hypothesis that IL-tolerant cellulases are also halotolerant is therefore confirmed.

Table 4.2: Traits of recently published cellulases including those investigated in this study, that were assayed in the presence of ILs

Cellulase	Specific activity (U/mg)	Relative activity (%) in (vol/vol %) IL	Residual activity (%) (incubation time) in (vol/vol %) IL	activity in NaCl	$T_{opt.}$ (°C)	Organism	Reference
Cel5A	294	65 in 20% [BMIM]Cl	80 (5 h) 40% [BMIM]Cl	50% in 3 M	75-80	*Thermoanaerobacter tengcongensis*	(Liang et al. 2011)
Celluclast® (Novozyme)	0.096	40 in 10% [MMIM][DMP]	40 (11 d) 10% [MMIM][DMP]	50% in 3 M	45	*Trichoderma reesei*	(Engel et al. 2010)
T. viride	11	40 in 5% [EMIM]Ac	0 (15 h) 15% [EMIM]Ac	-*	30	*Trichoderma viride*	(Datta et al. 2010)
Tma Cel5A	30	40 in 20% [EMIM]Ac	44 (15 h) 15% [EMIM]Ac	65% in 2 M	80	*Thermotoga maritima*	(Datta et al. 2010)
Pho EG	1.9	95 in 20% [EMIM]Ac	79 (15 h) 15% [EMIM]Ac	85% in 2 M	>95°C	*Pyrococcus horikoshii*	(Datta et al. 2010)
Cellulase preparation	-	-	40 (1 d) 50% [BMIM]Cl, but 0 (1 d) 30% [BMIM]Cl	-	-	*Penicillium janthinellum*	(Adsul 2009)
CelA10	2.4	74 in 30% [BMPL][OTF]	0.8 (17 h) 60% [BMPL][OTF]	50% in 4 M	55	metagenome	(Pottkämper et al. 2009)
CelA24	23.3	2 in 30% [BMPL][OTF]	0.02 (17 h) 60% [BMIM][OTF]	30% in 4 M	55	metagenome	(Pottkämper et al. 2009)
CelA2	11.9	54 in 30% [EMIM][OTF]	11 (5 d) 60% [BMPL][OTF]	80% in 4 M	55	metagenome	This work
CelA3	1.9	68 30% [EMIM][OTF]	79 (4 d) 60% [BMIM]Cl	50% in 4 M	70	metagenome	This work
CelA84	7.5	8 in 30% [BMPL][OTF]	81 (4 d) 60% [BMPL][OTF]	50% in 4 M	46	metagenome	This work
CelA	8.5	131 in 30% [EMIM][OTF]	115 (4 d) 60% [EMIM][OTF]	100% in 5 M	83	*T. maritima*	(Liebl et al. 1996)
Cel5K	8.1**	88 in 30% [EMIM][OTF]	138 (4 d) 60% [EMIM][OTF]	-	96	metagenome	Mientus, pers. comm

* no data available
** The $T_{opt.}$ of Cel5K was 96°C where the enzyme showed a specific activity of 78.9 U/mg, but the activity measurements in the presence of ILs were performed at 70°C, where the enzyme showed a specific activity of 8.1 U/mg.

On primary structure level halotolerance is often correlated with an elevated content of acidic aa and a lower proportion of basic aa. Therefore a hydrated salt ion network associated with the tertiary or quaternary structure of the protein occurs that stabilises the protein (Madern et al. 1995). In Fig. 4.1 the aa composition of the investigated cellulases is shown. An increased proportion of acidic aa could be detected for CelA and CelA2, with CelA2 marking the most halotolerant enzyme among the metagenomic cellulases, and CelA, whose activity was increased with the addition of 5 M NaCl at elevated temperatures (Liebl et al. 1996).

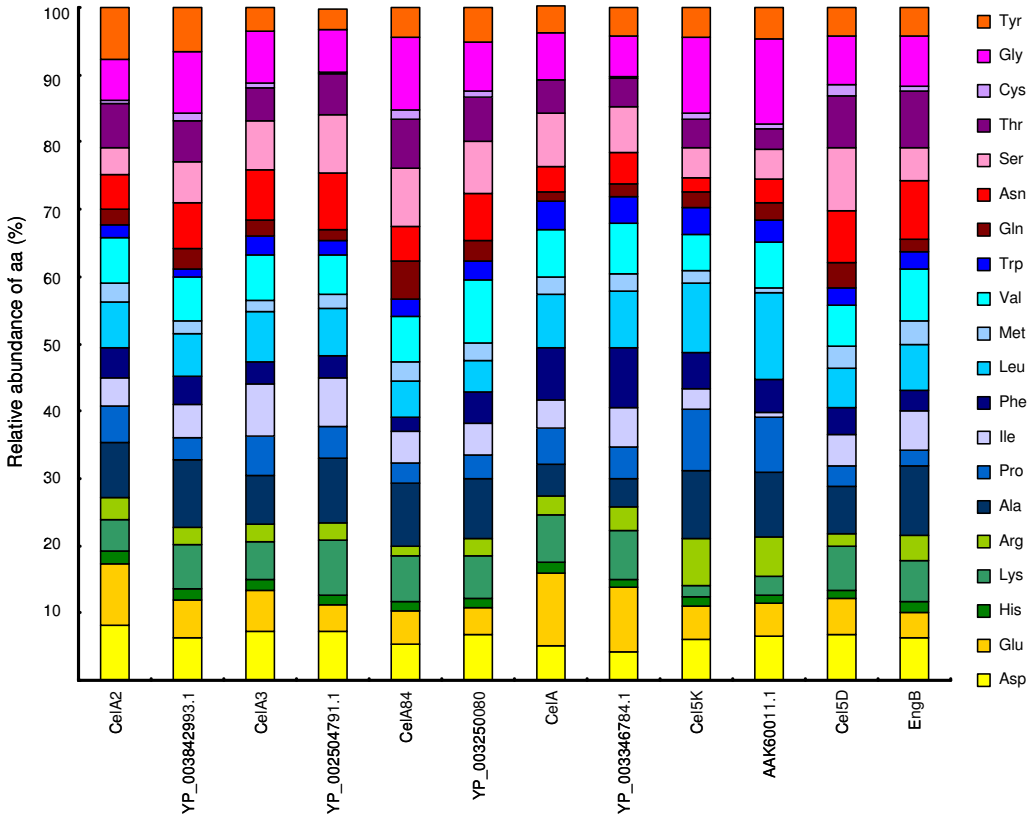

Figure 4.1: Relative abundance of the 20 proteinogenic aa in the investigated cellulases and in their respective closest relatives indicated by their Acc. No.

Amino acids illustrated in shades of red are polar, those in blue nonpolar, those in green basic and those in yellow acidic.

4.2.3 Thermostability of cellulases

The stability of proteins can be described on thermodynamic and on kinetic level. The thermodynamic stability depends on the equilibrium of unfolding and refolding rates which on their part depend on the Gibbs energy level of the native and the denatured state (Ogasahara et al. 1998). The kinetic stability is dependent on the activation energy for unfolding. Hyperthermophilic proteins have been described to have a very slow rate of unfolding arising from a high kinetic stability (Kaushik et al. 2002, Ogasahara et al. 1998).

It has been reported that thermostable enzymes generally show an enhanced tolerance against solvents and salts (Vazquez-Figueroa et al. 2008). The IL-tolerant enzymes investigated in this study shared the feature of moderate to high thermostability. CelA2 and CelA did not lose activity remarkably after 4 days at their respective optimum temperature. CelA84 was only moderately stable. After 4 days incubation at its temperature optimum (46°C) it retained 13% relative activity. Cel5K showed 65% relative activity after 3 days at 70°C and CelA3 exhibited 14% relative activity after 4 days at 70°C, but was stable at 55°C. The cellulase from *Thermoanaerobacter tengcongensis* MB4, that was active in the presence of 1 M [AMIM]Cl (approx. 16% (vol/vol)) and 1 M [BMIM]Cl (approx. 20% (vol/vol)) (54 and 65% relative activity, respectively), had a temperature optimum between 75 and 80°C (Liang et al. 2011). Also a hyperthermophilic endoglucanase from *Pyrococcus horikoshii*, with a $T_{opt.}$ higher than 95°C, lost only 5% activity in 20% (vol/vol) [EMIM]Ac (Datta et al. 2010). The data summarised in Table 4.2 indicate that thermostability or thermophilicity might be features linked to IL-tolerance.

Thermostability can be related to different features of proteins. These are e.g. a decrease in the thermolabile amino acids cysteine and aspartic acid (Unsworth et al. 2007). Furthermore metal ion binding, an elevated content of polar amino acids (resulting in a higher proportion of hydrogen bonds) and the reduction of surface loops have been described. Additionally stronger interactions like hydrophobic or aromatic interactions as well as ion pairs and salt bridges have been found in thermophilic proteins (Unsworth et al. 2007). Altogether, the intramolecular binding level is stronger which results in a tight packaging of the protein. Corresponding to the stronger binding level, the thermodynamic stability is higher for thermophilic than for mesophilic or psychrophilic enzymes and the unfolding rate is lower. The high energy level of intramolecular forces in

thermostable proteins also correlates with a rather rigid 3-dimensional structure at moderate temperatures which results in lower activity at these temperatures. The energy provided by elevated temperatures on the other hand results in a higher flexibility corresponding with higher reaction rates. The tight packaging results furthermore in the reduction of solvent exposed surface area.
Among the enzymes investigated in this study, an increase in polar amino acids could be found for CelA84. A decrease in cysteine residues could was detected for CelA and CelA2 while a decrease in aspartic acid is detected for CelA and CelA84 (Fig. 4.1).
A correlation between IL-tolerance and thermophilicity and thermostability has been foreshadowed before (Datta et al. 2010) and is reasonable because the solvent exposed surface area is reduced and because the activation energy for unfolding is higher in thermophilic proteins.

Altogether the data collected indicated that IL-tolerance correlates with thermostability or thermophilicity and also with halotolerance (Table 4.2). The modular structure and phylogenetic derivation on the other hand do not seem to influence IL-tolerance unless these are coherent with halotolerance or thermophilicity.

4.2.4 Activity and stability in the presence of ILs

The ILs used in this study were 1-butyl-3-methylimidazolium chloride ([BMIM]Cl), 1-butyl-3-methylimidazolium trifluoromethanesulfonate ([BMIM][OTF]), 1-butyl-2,3-dimethylimidazolium chloride ([BMMIM]Cl), 1-ethyl-3-methylimidazolium trifluoroacetate ([EMIM][ATF]), 1-butyl-1-methylpyrrolidinium trifluoromethanesulfonate ([BMPL][OTF]) and 1-ethyl-3-methylimidazolium trifluoromethanesulfonate ([EMIM][OTF]).

The activity of the investigated cellulases was assayed in the presence of 30% (vol/vol) IL at 37°C and at the respective optimum temperature (Cel5K was assayed at 70°C instead of 96°C); the relative activities are given in Fig. 4.2. The highest activity in the presence of 30% (vol/vol) was 11.15 U/mg, corresponding to 131% relative activity, observed for CelA in [EMIM][OTF] at 83°C. Also the other enzymes showed high activity, though CelA84 was

inactive in the presence of all ILs at its optimum temperature, 46°C, and showed furthermore the lowest relative activity in the presence of ILs at 37°C with maximal 11% in 30% (vol/vol) [EMIM][OTF]. CelA2, CelA84 and CelA were almost inactive in the presence of 30% (vol/vol) of the chloride containing ILs [BMIM]Cl and [BMMIM]Cl, but were active in the presence of the other ILs, e.g. 53% relative activity of CelA2 in [EMIM][OTF] at 55°C. In contrast, CelA3 exhibited comparable activity in the presence of all tested ILs. Compared to previously published enzymes, the relative activity of the cellulases in the presence of ILs was high. Up to now published cellulases showed 40% activity in 5% (vol/vol) [EMIM]Cl (Datta et al. 2010) or 2% in 30% (vol/vol) [BMPL][OTF], respectively (Pottkämper et al. 2009). In a study investigating metagenomic cellulases from soil and water samples nearly all enzymes lost their complete activity in the presence of 30% (vol/vol) IL (Pottkämper et al. 2009). The most active enzyme so far described is Cel5A, a GH5 endoglucanase from *Thermoanaerobacter tengcongensis* that showed 65% relative activity (191 U/mg) in 20% (vol/vol) [BMIM]Cl (Liang et al. 2011). The highest relative activity exhibited *Pho* EG from *Pyrococcus horikoshii* with 95% (1.8 U/mg) in 20% (vol/vol) [EMIM]Ac (Datta et al. 2010).

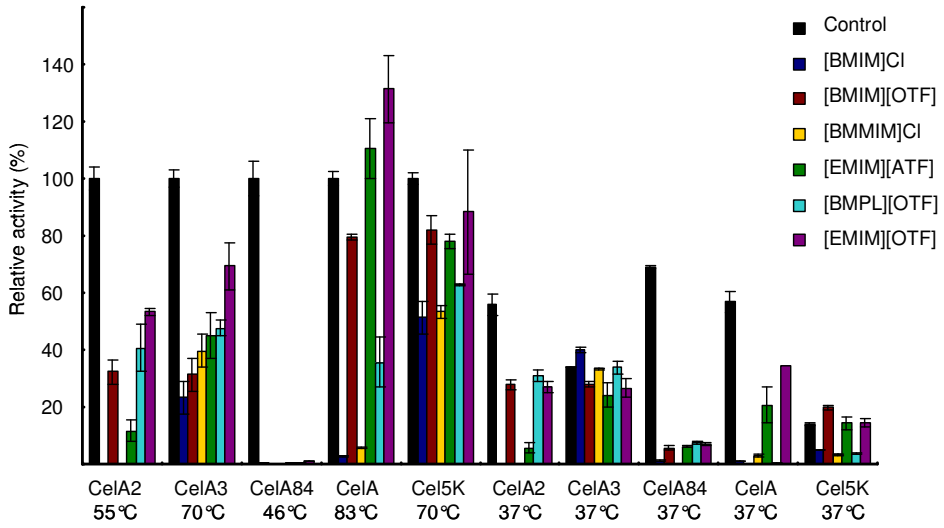

Figure 4.2: Relative activity of the investigated enzymes in the presence of 30% (vol/vol) IL at 37°C and at the respective temperature optimum (Cel5K was assayed at 70°C instead of 96°C)

The activity in the control at the higher temperature was set as 100% relative activity.

The enzymes' long term stability was tested in 60% (vol/vol) IL at room temperature whereby the activity was measured in 30% (vol/vol) IL at 37°C. In these assays CelA3 and Cel5K were the most stable enzymes. CelA3 did not lose activity after 4 days in the presence of 60% (vol/vol) [BMIM]Cl and remained active also in the presence of the ILs [BMIM][OTF] and [BMMIM]Cl. Cel5K stayed highly active in the presence of all ILs except [BMMIM]Cl and [EMIM][ATF]. CelA84 and CelA were stable in the presence of only one IL each, [BMPL][OTF] or [EMIM][OTF], respectively, while CelA2 was not stable in any of the tested ILs. The long term stability was furthermore investigated at the respective optimum temperature (Cel5K was assayed at 70°C instead of 96°C), whereby the activity measurement was also performed at $T_{opt.}$. Thereby none of the enzymes showed notable activity after a pre-incubation period of 1 to 3 days. In comparison with previously described enzymes the long term stability of the investigated enzymes was remarkably high. Except one enzyme preparation that has been tested after an incubation period of 11 days in the presence of ILs (Engel et al. 2010), the stability of cellulases in the presence of ILs was only assayed after short incubation periods (Table 4.2). Many enzymes lost most of their activity already after hours (Datta et al. 2010, Pottkämper et al. 2009).

The ILs used in this study are all water-miscible and therefore dissociate into the respective ions in aqueous solution as inorganic salts do. The activity of proteins in aqueous salt solutions can be roughly predicted along the rules of the Hofmeister series (Hofmeister 1888, Kunz et al. 2004). Therein ions are classified depending on their chaotropicity or kosmotropicity. Chaotropic ions are usually large in dimension and have a low charge density. Therefore they interact only weakly with water molecules and break the structure of water ("structure-breaker"). Chaotropic ions support the denaturation of proteins by lowering the hydrophobic effects and therefore by supporting the soluble unfolded form of the protein. Kosmotropic ions are small and strongly charged. All multivalent ions are highly hydrated and therefore kosmotropic. As their interaction with water is strong, they are called "structure-maker" (Zhao 2006). Kosmotropic ions induce higher hydrophobic effects in the protein and therefore stabilise its structure. Nevertheless, at higher concentrations of kosmotropic

salts, proteins precipitate as they are excluded from water ("salting-out"). For the quantification of kosmotropicity/chaotropicity different data and models have been established. The viscosity *B*-coefficient seems to correlate well with kosmotropicity (Zhao 2006). This coefficient describes the relative viscosity of a liquid as a function of the concentration of dissolved ions in correlation with ion radius and charge. For cations the hydration radius is furthermore taken into account and for anions the structural entropy (Zhao 2006). The kosmotropicity of ions also affects the water activity (Lo Nostro et al. 2005). This term describes the mass action effects of water on hydrolytic equilibria and therefore describes the intensity with which water associates with other compounds. Water activity and therefore water availability is reduced by the addition of salts or ILs (Halling 1994). As kosmotropes are highly hydrated, they bind more water and therefore reduce water activity to a greater extent than chaotropes. Water is an important component of many reactions. As the degradation of cellulose is a hydrolytic reaction, water participates directly in the reaction. It is furthermore important for the salvation of reaction products, for the stabilisation of transition states and for the flexibility of the enzyme (Halling 2004, Yang et al. 2004).

Many studies have demonstrated that kosmotropic anions and chaotropic cations usually stabilise proteins while kosmotropic cations and chaotropic anions tend to destabilise them (Zhao 2005). Equal tendencies have also been reported in studies using aqueous IL solutions (Maruyama et al. 2002, Park and Kazlauskas 2001). Nevertheless several enzymes have been shown to exhibit activity not in accordance with the rules of the Hofmeister series displaying individual requirements for solvent properties (Yang 2009, Yang et al. 2010).

The exact mechanism of how the kosmotropicity or chaotropicity of ions affect proteins is still not completely elucidated. Initially it was proposed that Hofmeister effects are mainly determined by the ions' impact on general water properties (Baldwin 1996, Zhang and Cremer 2006). Recent studies, however showed direct ion-macromolecule interactions and interactions of ions with the water molecules in the first hydration shell of the macromolecule (Zhang and Cremer 2006).

Regarding the ILs used in this study, chloride shows characteristics between kosmotropicity and chaotropicity and is therefore a so-called borderline ion (Zhao 2006). [OTF] is a chaotropic anion (Constantinescu et al. 2007) and [ATF] a kosmotropic one (Zhao 2006). Small organic cations are chaotropes while medium size organic ions are borderline ions and large ones are kosmotropes; [EMIM] was described as a weak chaotrope and [BMIM] as a weak kosmotrope or a borderline ion (Zhao 2006). For [BMMIM] no definition could be observed, but a rather kosmotropic character can be assumed as it is even larger than [BMIM]. [BMPL] has only been described as being more kosmotropic than [EMIM] and less than [BMIM] (Constantinescu et al. 2007). The ILs used in this study therefore represent a broad range of different ILs (Table 4.3).

Table 4.3: Kosmotropicity/chaotropicity of the ILs and the salt used in this study

IL/Salt	Cation (k/c)*	Anion (k/c)*
NaCl	k/c	k/c
[BMIM]Cl	k	k/c
[BMIM][OTF]	k	c
[BMMIM]Cl	k**	k/c
[EMIM][ATF]	c	k
[BMPL][OTF]	k/c***	c
[EMIM][OTF]	c	c

*k: kosmotrope, c: chaotrope, k/c: borderline ion; determined by viscosity *B*-coefficient and other methods referring to (Zhao 2006) and (Constantinescu et al. 2007)
**not described in literature but assumed to be a kosmotrope according to (Zhao 2006)
***only described as being located between [EMIM] and [BMIM] concerning kosmotropicity (Constantinescu et al. 2007)

According to the Hofmeister series, the highest cellulolytic activity should have been observed in the presence of [EMIM][ATF] as it is composed of a chaotropic cation and a kosmotropic anion, and in contrast the lowest in the presence of [BMIM][OTF] that consists of a kosmotropic cation and a chaotropic anion (Table 4.3). Though high activity was observed for all tested enzymes in the presence of [EMIM][ATF], the enzymes' activity in the presence of [BMIM][OTF] was similar to that in the presence of [EMIM][ATF] (Fig. 4.2), while the worst activity was usually recovered in the presence of [BMIM]Cl and

[BMMIM]Cl. Therefore Hofmeister effects do not seem to be decisive for the activity of the investigated enzymes in the presence of ILs.

Other properties of ILs have been taken into account to be affecting enzymes' activities. These are viscosity, H-bond basicity and nucleophilicity, polarity and hydrophobicity (Zhao et al. 2009).
In this study, the best activity was generally measured in the presence of [BMIM][OTF], [EMIM][ATF] and [EMIM][OTF] for all enzymes. These represent the least viscous of the tested ILs. It has been shown that the viscosity of the solvent can influence k_{cat} (Wood et al. 2009). The character of influence is determined by the rate determining factor of the reaction. If mass transfer is rate determining, the reaction is slowed down in highly viscous solvents. On the other hand viscosity is beneficial, if the stability of intermediates is rate determining. Furthermore it has been shown that the relaxation time increases with higher viscosity of the solvent (Chrunyk and Matthews 1990). Nevertheless high viscosity mainly impairs reaction rates but is not the determining factor in affecting enzyme activity (Zhao et al. 2009).

The lowest activity of all enzymes was usually observed in the presence of the chloride containing ILs, [BMIM]Cl and [BMMIM]Cl. CelA2, CelA84 and CelA showed almost no activity in their presence. It has been described that anions have a dominant effect over cations on enzyme behaviour (Kaar et al. 2003). In contrast, CelA2 and CelA84 showed 80% and 50% relative activity after 34 days in 4 M NaCl, respectively. CelA was even stabilised with the addition of 5 M NaCl at 85 to 95°C (CelA, (Liebl et al. 1996)). An alkaline phosphatase has been described to show highest activity in aqueous solutions containing salts whose ions have similar kosmotropic/chaotropic properties (Yang et al. 2010). This might be due to the effect that ions with similar kosmotropic/chaotropic properties and opposite charge tend to associate and therefore to affect the protein less (Yang et al. 2010). The high activity in the presence of NaCl could be explained with this hypothesis as Na^+ and chloride have very similar properties (Table 4.3).

H-bond basicity and nucleophilicity are two further features that can be important criterions for enzymes' activities in the presence of ILs (Yang et al. 2009). These characteristics are mainly determined by the anion. Among the anions investigated in this study, chloride is the most nucleophilic and H-bond basic anion (Henderson 2006, Henderson 2007, Linert et al. 1993). It has been shown furthermore that the nucleophilicity of chloride is influenced by the corresponding cation (Lancaster and Welton 2004). This explains the different behaviour of the investigated enzymes in the presence of NaCl and [BMIM]Cl as NaCl is a neutral salt while [BMIM]Cl is basic. High H-bond basicity and nucleophilicity of IL anions result in the formation of H-bonds with the protein leading to its denaturation (De Los Ríos et al. 2007, Zhao et al. 2008). Hence these features of the ILs [BMIM]Cl and [BMMIM]Cl possibly determine their negative influence on several enzymes' activities. On the other hand, high nucleophilicity and H-bond basicity are advantageous for the solvation of cellulose as the H-bond network of cellulose is broken by the formation of H-bonds between chloride and carbohydrate carboxylic protons; the substrate can therewith be dissolved, e.g. in [BMIM]Cl (Remsing et al. 2006). This results in the challenging problem to design ILs which dissolve cellulose but do not affect the structure and activity of cellulases.

Other solvent characteristics that have been described to affect enzyme activity are polarity and hydrophobicity (Gorman and Dordick 1992, Zhao et al. 2009). Several groups have described that rather nonpolar hydrophobic solvents have beneficial effects on enzyme activity as they do not strip the essential water off the enzyme (Gorman and Dordick 1992). Other groups, however, described that polarity and hydrophobicity did not affect enzyme performance (Narayan and Klibanov 1993) and also in hydrophilic ILs high enzyme activities could be observed (Park and Kazlauskas 2001). The ILs used in this study are all hydrophilic and the degree of polarity/hydrophilicity can be assumed to be in about the same range according to a previous survey (Reichardt 2005). The results of this study therefore confirm that high enzyme activities can be observed in the presence of hydrophilic ILs. On the other hand the enzymes' differing activities observed in this study in the presence of different ILs cannot be explained by the polarity/hydrophilicity of the used ILs.

Interestingly, none of the enzymes was active in the presence of [EMIM][ATF] after an incubation period of several days, though their direct activity in this IL was high. In contrast to the sensitivity towards the chloride containing ILs, that was shown by 3 out of the 5 investigated enzymes, [EMIM][ATF] affected all investigated cellulases. Furthermore the inactivation time was longer. This IL probably inactivates via a different mechanism than the chloride containing ILs do, though [ATF] is also moderately nucleophilic and the negative effect of [BMIM]Cl and [BMMIM]Cl can be assumed to be due to the high nucleophilicity of chloride. The reason for the instability of all enzymes in the presence of [EMIM][ATF] may furthermore be found in its anion, trifluoroacetate. Trifluoroacetic acid reacts with active hydrogen as such in hydroxy and amino groups and can therefore influence the protein structure (Yoon 2007).

The activity and long term stability of the investigated cellulases were lower at elevated temperatures. None of the enzymes showed activity after pre-incubation for 1 to 3 days at the respective optimum temperature. CelA84 did not even exhibit cellulolytic activity in the presence of any IL at its optimum temperature without pre-incubation. Furthermore the temperature optima determined in the presence of ILs were lower than those determined in buffer. This has been described previously for cellulases (Datta et al. 2010, Pottkämper et al. 2009, Yang et al. 2009). Thermal unfolding at lower temperatures can be explained because the intramolecular forces of cellulases are impaired by the presence of ILs. Therefore less energy has to be provided for unfolding. Furthermore, thermal unfolding at lower temperatures has been described for proteins in the presence of chaotropic reagents like guanidinium salts (Pace 1975).

On enzyme level the most noticeable observation is that in contrast to CelA2, CelA84 and CelA, CelA3 and Cel5K seemed to be rather generally IL-tolerant. While CelA2, CelA84 and CelA were almost inactive in the presence of the chloride containing ILs, CelA3 and Cel5K showed activity in the presence of all tested ILs. Additionally, CelA3 and Cel5K were much more long term stable in the presence of ILs than the other enzymes. Interestingly, CelA3 and Cel5K showed lower general activity at 37°C than the other enzymes. This lower

activity could be, beside other effects, explained by a rather rigid structure. This is due to high intramolecular forces that furthermore result in a higher thermodynamic stability. Consistently, the two enzymes with the lowest stability in the presence of ILs, CelA2 and CelA84, also have the lowest temperature optima, i.e. a lower energy need for unfolding. The inactivity of CelA84 at its optimum temperature in the presence of ILs can also be related to a low kinetic stability. The different behaviour of CelA2, CelA84 and CelA on the one hand and CelA3 and Cel5K on the other hand could not be linked to the phylogenetic derivation or explained with the modular structure of the enzymes or their aa sequence. CelA2, CelA84 and CelA are probably more sensitive to the nucleophilic attack of chloride in [BMIM]Cl and [BMMIM]Cl than the other enzymes. The reason therefore must be in their tertiary structure. To elucidate the different behaviour of IL-tolerant cellulases, crystallisation is a future prospect.

Regarding the results of this study it can be assumed that a combination of different effects determines the activity of individual enzymes in the presence of the respective ILs, whereupon on IL level nucleophilicity and viscosity seem to be more important than others. On enzyme level thermophilicity and the corresponding stability and the reduction of solvent exposed residues which are sensitive to a nucleophilic attack seem to be most important.

4.3 Industrial perspectives

Nowadays, bioethanol is usually produced by the hydrolysis of starch or smaller sugar chains from sugarcane or cereals (Akoh et al. 2008), which is in conflict with the food supply of people and cattle. In contrast, the generation of ethanol from cellulose does not directly affect food supply. Possible sources of cellulose are wood and different grasses as well as crop waste and forestry residues.

Different chemical methods for cellulose hydrolysis have been described. These are i.e. acid-catalysed hydrolysis, catalytic conversion into different products like Furan-based biofuels and processing in ILs (Van de Vyver et al. 2011). The chemical processes have the disadvantage that the energy need is

usually high and that toxic waste occurs that has to be depolluted cost-intensively. On the other hand the enzymatic process fails as the substrate is insoluble in water and the enzymes need an aqueous medium for efficient hydrolysis. Some ILs dissolve cellulose (Heinze et al. 2005, Swatloski et al. 2002) and would therefore provide an alternative reaction medium if cellulases were active and stable in their presence. A major disadvantage of the application of ILs is their expensiveness (e.g. 100 g [BMIM]Cl cost 128 €, Roth, Karlsruhe, Germany). To recycle the ILs does not effectively circumvent this problem as it is a challenging technological task and furthermore by-products occur (Shill et al. 2011). Altogether it is not yet elucidated, if it is possible to generate bioethanol from cellulose in an economic way (Chen and Qiu 2010).

For the complete conversion of cellulose to glucose three types of enzymes are necessary. These are endoglucanases which cut cellulose anywhere at amorphous regions of the cellulose strand, exoglucanases that liberate glucose-oligomers (usually the disaccharide cellobiose) from the end of the cellulose strand and cellobiases that hydrolyse cellobiose to glucose. All enzymes discovered in this study were screened using CMC as a substrate. This is a type of cellulose that does not occur in nature but that is easy to handle in the laboratory as it is amorphous and soluble in water. In this study only endoglucanases have been detected. This is the group of enzymes usually detected by screening on CMC (Pottkämper et al. 2009, Voget et al. 2006). Cellobiases hydrolyse the disaccharide cellobiose to glucose and are usually not able to hydrolyse cellulose and therefore could not be detected within this assay. Exoglucanases, however, liberate glucose oligomers from the end of the cellulose strand. They often rather act on crystalline cellulose (Zhang et al. 2010) but could have been detected in this study.
None of the enzymes was, accessorily to the activity against CMC, able to hydrolyse crystalline cellulose. Consequentially, an additional exoglucanase which hydrolyses the crystalline substrate and an enzyme that hydrolyses cellobiose to glucose would be necessary for a complete reaction system. Nevertheless, the investigated cellulases showed remarkable activity and stability in the presence of ILs and could therefore be interesting candidates for an industrial application in combination with exoglucanases and cellobiases.

Among the tested enzymes only CelA and Cel5K were capable of hydrolysing crystalline cellulose that has been dissolved in 1-butyl-3-methylimidazolium acetate ([BMIM]Ac) or 1-ethyl-3-methylimidazolium acetate ([EMIM]Ac) and afterwards precipitated with water. This is not self-evident as CelA3, that was inactive against this substrate, was the only enzyme containing a CBM, that is essential for the hydrolysis of crystalline cellulose (Bolam et al. 1998, Carvalho et al. 2004). Nevertheless not all CBMs mediate the attachment to crystalline cellulose. CelA3 contained two CBM family 17_28 domains which bind to amorphous cellulose (Boraston et al. 2003). Altogether CelA and Cel5K had the highest affinity to rather crystalline substrates though none of them contained a CBM. Via the treatment with ILs, crystalline cellulose becomes more amorphous and is therefore less recalcitrant to enzymatic hydrolysis (Li et al. 2010). Therefore an exoglucanase would be unnecessary in this kind of reaction system and the enzymes would only need to tolerate smaller amounts of ILs. But, disadvantageously, the required reaction volume is higher.

Furthermore crude extracts of the clones pFosCelA2, pFosCelA3 and pFosCelA84 were remarkably active in the presence of ethanol (38-66% relative activity in 20% (vol/vol) ethanol), the possible end product of the process. This would be advantageous, if a one-pot process (simultaneous saccharification and fermentation, SSF) was aimed. It has been shown, that ethanol inhibits cellulases (Philippidis et al. 1993, Podkaminer et al. 2011, Wu and Lee 1997). E. g. the cellulase from *Trichoderma reesei* showed 50% activity in 7% ethanol (Philippidis et al. 1993). Furthermore it has been shown that with increasing temperature ethanol has a stronger inhibitory effect (Podkaminer et al. 2011, Wu and Lee 1997) which has been described for ILs in this study as well.

The crucial property is the enzymes' activity and long term stability in aqueous IL solutions. Among the investigated cellulases CelA3 and Cel5K seemed to be the most robust and most stable enzymes, whereby Cel5K was additionally able to hydrolyse the IL dissolved and with water precipitated crystalline cellulose. Furthermore these enzymes' temperature optima were 70°C and

96 °C respectively, and would therefore allow a reaction at moderate temperatures what would be advantageous with respect to the energy costs.

4.4 Conclusion & outlook

Within this study it has been shown that metagenomics is an appropriate tool for the discovery of novel IL-tolerant cellulases. Altogether 14 cellulases have been discovered of which some showed remarkable activity and long term stability in the presence of different ILs. In comparison with further enzymes it has been elucidated that halotolerance and thermostability generally seem to correlate with IL-tolerance while ORF structure and phylogeny do not, unless these are coherent with other enzyme characteristics linked to IL-tolerance. Among the properties of ILs, H-bond basicity and viscosity seem to be determinant for enzyme activity. To elucidate the individual activities of cellulases in the presence of different ILs, crystallisation is a future prospect.
Applying these results it is possible to identify further, even more IL-tolerant endoglucanases. For the completion of the cellulose hydrolysis system, cellobiases and exoglucanases that are active and stable in the presence of ILs can be searched for in a more directed manner. Furthermore directed evolution of the discovered endoglucanases is an opportunity to optimise the enzymes' properties. To test the applicability of the enzymes in a biotechnological process should be a following step.

5 Abstract

Cellulose is a ubiquitous biopolymer consisting of β-1,4-linked D-glucose that can be used as a source for glucose and subsequent molecules, e.g. ethanol. The enzymatic hydrolysis has one major drawback: Crystalline cellulose is not soluble in water and the enzymes need an aquatic medium for efficient hydrolysis. Some ionic liquids (ILs) have been described to dissolve cellulose and would therefore provide an alternative reaction medium, if cellulases were functional in the presence of ILs.

The aim of this work was to discover novel metagenomic cellulases that are active and stable in the presence of ILs. Therefore hydrolytic habitats were evaluated on activity and biodiversity level. Metagenomic libraries derived from the most auspicious habitats, elephant faeces (20,000 clones), a biogas plant sample (9,600 clones) and an enrichment culture inoculated with the shipworm *Teredo navalis* (3,600 clones), were screened with a function based approach. Altogether 14 cellulolytic clones were identified. The most active clones were sequenced via pyrosequencing and were evaluated for their performance in the presence of 30% (vol/vol) IL and after overnight incubation in 60% (vol/vol) IL. Though many of the clones lost almost their complete activity in the presence of the majority of ILs used in this study, two clones from the library constructed with DNA derived from the biogas plant (pFosCelA2 and pFosCelA3) and several clones from the library constructed with DNA derived from elephant faeces (among others pFosCelA84) were active and stable in the presence of ILs. Therefore the respective enzymes were purified and furthermore investigated. CelA2 showed highest identity (41%) to a glycoside hydrolase (GH) family 9 protein from *Clostridium cellulovorans* and comprised an additional Ig-like domain. CelA3 was closest related (60% identity) to a GH5 cellulase from *Clostridium cellulolyticum* and included a carbohydrate binding module while CelA84 did not comprise any additional domain and was 41% identical to a GH5 enzyme from *Fibrobacter succinogenes*. In addition to their IL-tolerance, all three enzymes shared the features of moderate to high thermostability and high halotolerance. For this reason, the enzymes were compared with two further thermostable cellulases; CelA was a GH12 endoglucanase from *Thermotoga maritima* and Cel5K was a metagenomic

cellulase derived from the Avachinsky Crater in Kamchatka with highest identity (61%) to a GH5 enzyme from *Thermus caldophilus*. Both enzymes did not comprise any additional domain.

CelA3 showed comparable activity in the presence of 30% (vol/vol) of all tested ILs (about 1 U/mg at 37°C and 1-2 U/mg at 70°C). Cel5K also showed activity in the presence of all tested ILs. The highest activity was observed in the presence of 30% (vol/vol) [EMIM][OTF] at 70°C (7.1 U/mg). CelA2, CelA84 and CelA were almost inactive in the chloride containing ILs [BMIM]Cl and [BMMIM]Cl though they were active in the presence of the other ILs. CelA84 showed about 0.5 U/mg in the other tested ILs at 37°C while CelA2 exhibited 6.4 U/mg in [EMIM][OTF] at 55°C. CelA exhibited with 11.4 U/mg in 30% (vol/vol) [EMIM][OTF] at 83°C the highest overall activity.

Some of the cellulases also exhibited very high long term stability in the presence of ILs that has never been described before. At room temperature, CelA3 was stable over 4 days in the presence of 60% (vol/vol) of several ILs ([BMIM]Cl, [BMIM][OTF] and [BMMIM]Cl). Cel5K did not lose any activity after 4 days in 60% (vol/vol) [BMIM][OTF] and [EMIM][OTF]. In contrast, the three enzymes that were inactive in the presence of 30% (vol/vol) [BMIM]Cl and [BMMIM]Cl exhibited considerably lower long term stability in the presence of ILs. At the respective optimum temperatures, none of the enzymes was stable in the presence of any of the tested ILs after a pre-incubation period of 1 to 3 days.

With the characterisation of the described cellulases it was shown that (A) metagenomics is an appropriate tool for the discovery of novel cellulases which are (B) very active and (C) long term stable in the presence of high IL concentrations, that (D) IL-tolerance depends on the reaction temperature, (E) halotolerance is linked to IL-tolerance, (F) thermostability is linked to IL-tolerance, (G) the phylogenetic derivation of cellulases does not correlate with IL-tolerance unless it is coherent with other properties related to IL-tolerance, (H) additional domains like carbohydrate binding modules and Ig-like domains are not necessary for IL-tolerance and that (I) each enzyme has to be investigated individually, as each enzyme has individual requirements for solvent properties.

6 References

Adsul, M. G., Terwadkar, A. P., Varma, A. J., Gokhale, D. 2009. Cellulases from *Penicillium janthinellum* mutants: solid-state production and their stability in ionic liquids. *BioResources* 4: 1670-1681.

Akoh, C. C., S. W. Chang, G. C. Lee, and J. F. Shaw. 2008. Biocatalysis for the production of industrial products and functional foods from rice and other agricultural produce. *J Agric Food Chem* 56: 10445-51.

Allgaier, M., A. Reddy, J. I. Park, N. Ivanova, P. D'Haeseleer, S. Lowry, R. Sapra, T. C. Hazen, B. A. Simmons, J. S. VanderGheynst, and P. Hugenholtz. 2010. Targeted Discovery of Glycoside Hydrolases from a Switchgrass-Adapted Compost Community. *PLoS ONE* 5: e8812.

Amann, R. I., W. Ludwig, and K. H. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59: 143-169.

Anbarasan, S., J. Janis, M. Paloheimo, M. Laitaoja, M. Vuolanto, J. Karimaki, P. Vainiotalo, M. Leisola, and O. Turunen. 2010. Effect of Glycosylation and Additional Domains on the Thermostability of a Family 10 Xylanase Produced by *Thermopolyspora flexuosa*. *Appl. Environ. Microbiol.* 76: 356-360.

Angelov, A., M. Mientus, S. Liebl, and W. Liebl. 2009. A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Systematic and Applied Microbiology* 32: 177.

Araki, Y., S. Karita, A. Tanaka, M. Kondo, and M. Goto. 2009. Characterization of Family 17 and Family 28 Carbohydrate-Binding Modules from *Clostridium josui* Cel5A. *Bioscience, Biotechnology, and Biochemistry* 73: 1028.

Baldwin, R. L. 1996. How Hofmeister ion interactions affect protein stability. *Biophys J* 71: 2056-63.

Baneyx, F. 1999. Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* 10: 411.

Bhat, M. K. 2000. Cellulases and related enzymes in biotechnology. *Biotechnology Advances* 18: 355.

Bolam, D. N., A. Ciruela, S. McQueen-Mason, P. Simpson, M. P. Williamson, J. E. Rixon, A. Boraston, G. P. Hazlewood, and H. J. Gilbert. 1998.

Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J* 331 (Pt 3): 775-81.

Boraston, A. B., D. N. Bolam, H. J. Gilbert, and G. J. Davies. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382: 769-81.

Boraston, A. B., E. Kwan, P. Chiu, R. A. J. Warren, and D. G. Kilburn. 2003. Recognition and Hydrolysis of Noncrystalline Cellulose. *Journal of Biological Chemistry* 278: 6120-6127.

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-54.

Bridgwater, A. V., and G. V. C. Peacocke. 2000. Fast pyrolysis processes for biomass. *Renewable and Sustainable Energy Reviews* 4: 1.

Carvalho, A. L., A. Goyal, J. A. Prates, D. N. Bolam, H. J. Gilbert, V. M. Pires, L. M. Ferreira, A. Planas, M. J. Romao, and C. M. Fontes. 2004. The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates beta-1,4- and beta-1,3-1,4-mixed linked glucans at a single binding site. *J Biol Chem* 279: 34785-93.

Chen, H., and W. Qiu. 2010. Key technologies for bioethanol production from lignocellulose. *Biotechnology Advances* 28: 556.

Chester, N., and D. R. Marshak. 1993. Dimethyl Sulfoxide-Mediated Primer Tm Reduction: A Method for Analyzing the Role of Renaturation Temperature in the Polymerase Chain Reaction. *Analytical Biochemistry* 209: 284.

Chrunyk, B. A., and C. R. Matthews. 1990. Role of diffusion in the folding of the alpha subunit of tryptophan synthase from *Escherichia coli*. *Biochemistry* 29: 2149.

Constantinescu, D., H. Weingärtner, and C. Herrmann. 2007. Protein Denaturation by Ionic Liquids and the Hofmeister Series: A Case Study of Aqueous Solutions of Ribonuclease A. *Angewandte Chemie International Edition* 46: 8887.

Daniel, R. 2004. The soil metagenome--a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* 15: 199-204.

Datta, S., B. Holmes, J. I. Park, Z. Chen, D. C. Dibble, M. Hadi, H. W. Blanch, B. A. Simmons, and R. Sapra. 2010. Ionic liquid tolerant

hyperthermophilic cellulases for biomass pretreatment and hydrolysis. *Green Chemistry* 12: 338.

De Los Ríos, A. P., F. J. Hernández-Fernández, F. A. Martínez, M. Rubio, and G. Víllora. 2007. The effect of ionic liquid media on activity, selectivity and stability of *Candida antarctica* lipase B in transesterification reactions. *Biocatalysis and Biotransformation* 25: 151-156.

Distel, D. L., E. F. DeLong, and J. B. Waterbury. 1991. Phylogenetic characterization and in situ localization of the bacterial symbiont of shipworms (*Teredinidae*: *Bivalvia*) by using 16S rRNA sequence analysis and oligodeoxynucleotide probe hybridization. *Appl. Environ. Microbiol.* 57: 2376-2382.

Doi, R. H., A. Kosugi, K. Murashima, Y. Tamaru, and S. O. Han. 2003. Cellulosomes from Mesophilic Bacteria. *J. Bacteriol.* 185: 5907-5914.

Doi, R. H., J.-S. Park, C.-c. Liu, L. M. Malburg Jr, Y. Tamaru, A. Ichiishi, and A. Ibrahim. 1998. Cellulosome and noncellulosomal cellulases of *Clostridium cellulovorans*. *Extremophiles* 2: 53.

Doi, R. H., and Y. Tamaru. 2001. The *Clostridium cellulovorans* cellulosome: An enzyme complex with plant cell wall degrading activity. *The Chemical Record* 1: 24.

Dowd, S., T. Callaway, R. Wolcott, Y. Sun, T. McKeehan, R. Hagevoort, and T. Edrington. 2008. Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiology* 8: 125.

Duan, C. J., L. Xian, G. C. Zhao, Y. Feng, H. Pang, X. L. Bai, J. L. Tang, Q. S. Ma, and J. X. Feng. 2009. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *Journal of Applied Microbiology* 107: 245.

Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman. 2005. Diversity of the Human Intestinal Microbial Flora. *Science* 308: 1635-1638.

Elend, C., C. Schmeisser, H. Hoebenreich, H. L. Steele, and W. R. Streit. 2007. Isolation and characterization of a metagenome-derived and cold-active lipase with high stereospecificity for (R)-ibuprofen esters. *J Biotechnol* 130: 370-7.

Elend, C., C. Schmeisser, C. Leggewie, P. Babiak, J. D. Carballeira, H. L. Steele, J. L. Reymond, K. E. Jaeger, and W. R. Streit. 2006. Isolation

and biochemical characterization of two novel metagenome-derived esterases. *Appl Environ Microbiol* 72: 3637-45.

Engel, P., R. Mladenov, H. Wulfhorst, G. Jager, and A. C. Spiess. 2010. Point by point analysis: how ionic liquid affects the enzymatic hydrolysis of native and modified cellulose. *Green Chemistry* 12: 1959.

Feng, Y., C. J. Duan, H. Pang, X. C. Mo, C. F. Wu, Y. Yu, Y. L. Hu, J. Wei, J. L. Tang, and J. X. Feng. 2007. Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Appl Microbiol Biotechnol* 75: 319-28.

Fontes, C. M. G. A., and H. J. Gilbert. 2010. Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. *Annual Review of Biochemistry* 79: 655-681.

Gabor, E. M., W. B. Alkema, and D. B. Janssen. 2004a. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* 6: 879-86.

Gabor, E. M., E. J. de Vries, and D. B. Janssen. 2004b. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environmental Microbiology* 6: 948-958.

Ganske, F., and U. T. Bornscheuer. 2005. Lipase-Catalyzed Glucose Fatty Acid Ester Synthesis in Ionic Liquids. *Organic Letters* 7: 3097.

Gillespie, D. E., S. F. Brady, A. D. Bettermann, N. P. Cianciotto, M. R. Liles, M. R. Rondon, J. Clardy, R. M. Goodman, and J. Handelsman. 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* 68: 4301-6.

Gorman, L. A., and J. S. Dordick. 1992. Organic solvents strip water off enzymes. *Biotechnol Bioeng* 39: 392-7.

Halling, P. J. 1994. Thermodynamic predictions for biocatalysis in nonconventional media: Theory, tests, and recommendations for experimental design and analysis. *Enzyme and Microbial Technology* 16: 178.

Halling, P. J. 2004. What can we learn by studying enzymes in non-aqueous media? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359: 1287-1297.

Heinze, T., K. Schwikal, and S. Barthel. 2005. Ionic liquids as reaction medium in cellulose functionalization. *Macromol Biosci* 5: 520-5.

Henderson, W. A. 2006. Glyme-Lithium Salt Phase Behavior. *The Journal of Physical Chemistry B* 110: 13177.

Henderson, W. A. 2007. Crystallization Kinetics of Glyme-LiX and PEO-LiX Polymer Electrolytes. *Macromolecules* 40: 4963.

Herrera, A., and C. S. Cockell. 2007. Exploring microbial diversity in volcanic environments: A review of methods in DNA extraction. *Journal of Microbiological Methods* 70: 1.

Hofmeister, F. 1888. Zur Lehre der Wirkung der Salze. Zweite Mittheilung. *Arch. Exp. Pathol. Pharmakol.* 24: 247-260.

Hu, Y., G. Zhang, A. Li, J. Chen, and L. Ma. 2008. Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. *Appl Microbiol Biotechnol* 80: 823-30.

Ito, S. 1997. Alkaline cellulases from alkaliphilic *Bacillus*: Enzymatic properties, genetics, and application to detergents. *Extremophiles* 1: 61.

Juretschko, S., G. Timmermann, M. Schmid, K.-H. Schleifer, A. Pommerening-Roser, H.-P. Koops, and M. Wagner. 1998. Combined Molecular and Conventional Analyses of Nitrifying Bacterium Diversity in Activated Sludge: *Nitrosococcus mobilis* and *Nitrospira*-Like Bacteria as Dominant Populations. *Appl. Environ. Microbiol.* 64: 3042-3051.

Kaar, J. L., A. M. Jesionowski, J. A. Berberich, R. Moulton, and A. J. Russell. 2003. Impact of Ionic Liquid Physical Properties on Lipase Activity and Stability. *Journal of the American Chemical Society* 125: 4125.

Kane, M. D., L. K. Poulsen, and D. A. Stahl. 1993. Monitoring the enrichment and isolation of sulfate-reducing bacteria by using oligonucleotide hybridization probes designed from environmentally derived 16S rRNA sequences. *Appl Environ Microbiol* 59: 682-6.

Kataeva, I., X. L. Li, H. Chen, S. K. Choi, and L. G. Ljungdahl. 1999. Cloning and sequence analysis of a new cellulase gene encoding CelK, a major cellulosome component of *Clostridium thermocellum*: evidence for gene duplication and recombination. *J Bacteriol* 181: 5288-95.

Kataeva, I. A., R. D. Seidel, 3rd, A. Shah, L. T. West, X. L. Li, and L. G. Ljungdahl. 2002. The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. *Appl Environ Microbiol* 68: 4292-300.

Kataeva, I. A., V. N. Uversky, J. M. Brewer, F. Schubot, J. P. Rose, B. C. Wang, and L. G. Ljungdahl. 2004. Interactions between immunoglobulin-like and catalytic modules in *Clostridium thermocellum* cellulosomal cellobiohydrolase CbhA. *Protein Eng Des Sel* 17: 759-69.

Katsoura, M. H., A. C. Polydera, L. Tsironis, A. D. Tselepis, and H. Stamatis. 2006. Use of ionic liquids as media for the biocatalytic preparation of flavonoid derivatives with antioxidant potency. *Journal of Biotechnology* 123: 491.

Kaushik, J. K., K. Ogasahara, and K. Yutani. 2002. The Unusually Slow Relaxation Kinetics of the Folding-unfolding of Pyrrolidone Carboxyl Peptidase from a Hyperthermophile, *Pyrococcus furiosus*. *Journal of Molecular Biology* 316: 991.

Kilpelainen, I., H. Xie, A. King, M. Granstrom, S. Heikkinen, and D. S. Argyropoulos. 2007. Dissolution of wood in ionic liquids. *J Agric Food Chem* 55: 9142-8.

Kim, S. J., C. M. Lee, B. R. Han, M. Y. Kim, Y. S. Yeo, S. H. Yoon, B. S. Koo, and H. K. Jun. 2008. Characterization of a gene encoding cellulase from uncultured soil bacteria. *FEMS Microbiol Lett* 282: 44-51.

Knietsch, A., S. Bowien, G. Whited, G. Gottschalk, and R. Daniel. 2003a. Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl Environ Microbiol* 69: 3048-60.

Knietsch, A., T. Waschkowitz, S. Bowien, A. Henne, and R. Daniel. 2003b. Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol* 69: 1408-16.

Kobayashi, Y., T. Shinkai, and S. Koike. 2008. Ecological and physiological characterization shows that *Fibrobacter succinogenes* is important in rumen fiber digestion — Review. *Folia Microbiologica* 53: 195.

Koivula, A., T. Reinikainen, L. Ruohonen, A. Valkeajärvi, M. Claeyssens, O. Teleman, G. J. Kleywegt, M. Szardenings, J. Rouvinen, T. A. Jones, and T. T. Teeri. 1996. The active site of *Trichoderma reesei* cellobiohydrolase II: the role of tyrosine 169. *Protein Engineering* 9: 691-699.

Kragl, U., M. Eckstein, and N. Kaftzik. 2002. Enzyme catalysis in ionic liquids. *Current Opinion in Biotechnology* 13: 565.

Kunz, W., J. Henle, and B. W. Ninham. 2004. 'Zur Lehre von der Wirkung der Salze' (about the science of the effect of salts): Franz Hofmeister's historical papers. *Current Opinion in Colloid & Interface Science* 9: 19.

Lancaster, N. L., and T. Welton. 2004. Nucleophilicity in Ionic Liquids. 3.1 Anion Effects on Halide Nucleophilicity in a Series of 1-Butyl-3-methylimidazolium Ionic Liquids. *The Journal of Organic Chemistry* 69: 5986.

Lee, S. H., T. V. Doherty, R. J. Linhardt, and J. S. Dordick. 2009. Ionic liquid-mediated selective extraction of lignin from wood leading to enhanced enzymatic cellulose hydrolysis. *Biotechnology and Bioengineering* 102: 1368.

Li, C., B. Knierim, C. Manisseri, R. Arora, H. V. Scheller, M. Auer, K. P. Vogel, B. A. Simmons, and S. Singh. 2010. Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. *Bioresource Technology* 101, 4900-4906.

Li, Y., M. Wexler, D. J. Richardson, P. L. Bond, and A. W. Johnston. 2005. Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned trp genes. *Environ Microbiol* 7: 1927-36.

Liang, C., Y. Xue, M. Fioroni, F. Rodriguez-Ropero, C. Zhou, U. Schwaneberg, and Y. Ma. 2011. Cloning and characterization of a thermostable and halo-tolerant endoglucanase from *Thermoanaerobacter tengcongensis* MB4. *Appl Microbiol Biotechnol*. 89: 315-325.

Liebl, W., P. Ruile, K. Bronnenmeier, K. Riedel, F. Lottspeich, and I. Greif. 1996. Analysis of a *Thermotoga maritima* DNA fragment encoding two similar thermostable cellulases, CelA and CelB, and characterization of the recombinant enzymes. *Microbiology* 142: 2533-2542.

Linert, W., R. F. Jameson, and A. Taha. 1993. Donor numbers of anions in solution: the use of solvatochromic Lewis acid-base indicators. *Journal of the Chemical Society, Dalton Transactions*: 3181.

Liu, J., W.-d. Liu, X.-l. Zhao, W.-j. Shen, H. Cao, and Z.-l. Cui. 2011. Cloning and functional characterization of a novel endo-β-1,4-glucanase gene from a soil-derived metagenomic library. *Applied Microbiology and Biotechnology* 89: 1083.

Lo Nostro, P., B. W. Ninham, A. Lo Nostro, G. Pesavento, L. Fratoni, and P. Baglioni. 2005. Specific ion effects on the growth rates of *Staphylococcus aureus* and *Pseudomonas aeruginosa*. *Physical Biology* 2: 1.

Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.-H. Schleifer. 2004. ARB: a software environment for sequence data. *Nucl. Acids Res.* 32: 1363-1371.

Lynd, L. R., P. J. Weimer, W. H. van Zyl, and I. S. Pretorius. 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66: 506-77, table of contents.

Madern, D., C. Pfister, and G. Zaccai. 1995. Mutation at a single acidic amino acid enhances the halophilic behaviour of malate dehydrogenase from *Haloarcula marismortui* in physiological salts. *Eur J Biochem* 230: 1088-95.

Makrides, S. C. 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev* 60: 512-38.

Martinez, A., S. J. Kolvek, C. L. Yip, J. Hopke, K. A. Brown, I. A. MacNeil, and M. S. Osburne. 2004. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Applied and Environmental Microbiology* 70: 2452-2463.

Maruyama, T., S. Nagasawa, and M. Goto. 2002. Poly(ethylene glycol)-lipase complex that is catalytically active for alcoholysis reactions in ionic liquids. *Biotechnology Letters* 24: 1341.

Meske, D. 2009. Identification of novel cellulases with biotechnological potential and comparative analysis of epiphytic bacteria on macroalgal surfaces. *Dissertation, Christian Albrechts Universität zu Kiel*.

Mitsumori, M., H. Minato, T. Sekizaki, I. Uchida, and H. Ito. 1996. Cloning, nucleotide sequence and expression of the gene encoding the cellulose-binding protein 1 (CBP1) of *Fibrobacter succinogenes* S85. *FEMS Microbiol Lett* 139: 43-50.

Murashima, K., C. L. Chen, A. Kosugi, Y. Tamaru, R. H. Doi, and S. L. Wong. 2002. Heterologous production of *Clostridium cellulovorans engB*, using protease-deficient *Bacillus subtilis*, and preparation of active recombinant cellulosomes. *Journal of Bacteriology* 184: 76-81.

Muyzer, G., E. C. de Waal, and A. G. Uitterlinden. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59: 695-700.

Narayan, V. S., and A. M. Klibanov. 1993. Are water-immiscibility and apolarity of the solvent relevant to enzyme efficiency? *Biotechnol Bioeng* 41: 390-3.

Ogasahara, K., M. Nakamura, S. Nakura, S. Tsunasawa, I. Kato, T. Yoshimoto, and K. Yutani. 1998. The Unusually Slow Unfolding Rate Causes the High Stability of Pyrrolidone Carboxyl Peptidase from a Hyperthermophile, *Pyrococcus furiosus*: Equilibrium and Kinetic Studies of Guanidine Hydrochloride-Induced Unfolding and Refolding. *Biochemistry* 37: 17537.

Pace, C. N. 1975. The stability of globular proteins. *CRC Crit Rev Biochem* 3: 1-43.

Pang, H., P. Zhang, C. J. Duan, X. C. Mo, J. L. Tang, and J. X. Feng. 2009. Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Curr Microbiol* 58: 404-8.

Park, H. J., J. H. Jeon, S. G. Kang, J. H. Lee, S. A. Lee, and H. K. Kim. 2007. Functional expression and refolding of new alkaline esterase, EM2L8 from deep-sea sediment metagenome. *Protein Expr Purif* 52: 340-7.

Park, S., and R. J. Kazlauskas. 2001. Improved Preparation and Use of Room-Temperature Ionic Liquids in Lipase-Catalyzed Enantio- and Regioselective Acylations. *The Journal of Organic Chemistry* 66: 8395.

Pereira, J. H., R. Sapra, J. V. Volponi, C. L. Kozina, B. Simmons, and P. D. Adams. 2009. Structure of endoglucanase Cel9A from the

thermoacidophilic *Alicyclobacillus acidocaldarius*. *Acta Crystallogr D Biol Crystallogr* 65: 744-50.

Philippidis, G. P., T. K. Smith, and C. E. Wyman. 1993. Study of the enzymatic hydrolysis of cellulose for production of fuel ethanol by the simultaneous saccharification and fermentation process. *Biotechnology and Bioengineering* 41: 846.

Podkaminer, K. K., X. Shao, D. A. Hogsett, and L. R. Lynd. 2011. Enzyme inactivation by ethanol and development of a kinetic model for thermophilic simultaneous saccharification and fermentation at 50 °C with *Thermoanaerobacterium saccharolyticum* ALK2. *Biotechnology and Bioengineering* 108: 1268-1278.

Pohar, A., I. Plazl, and P. Znidarsic-Plazl. 2009. Lipase-catalyzed synthesis of isoamyl acetate in an ionic liquid/n-heptane two-phase system at the microreactor scale. *Lab on a Chip* 9: 3385-3390.

Pottkämper, J. 2009. Cellulase-Aktivitäten nicht-kultivierter Mikroorganismen in ionischen Flüssigkeiten. *Dissertation, Universität Hamburg*.

Pottkämper, J., P. Barthen, N. Ilmberger, U. Schwaneberg, A. Schenk, M. Schulte, N. Ignatiev, and W. R. Streit. 2009. Applying metagenomics for the identification of bacterial cellulases that are stable in ionic liquids. *Green Chemistry* 11: 957-965.

Reichardt, C. 2005. Polarity of ionic liquids determined empirically by means of solvatochromic pyridinium N-phenolate betaine dyes. *Green Chemistry* 7: 339.

Remsing, R. C., R. P. Swatloski, R. D. Rogers, and G. Moyna. 2006. Mechanism of cellulose dissolution in the ionic liquid 1-n-butyl-3-methylimidazolium chloride: a 13C and 35/37Cl NMR relaxation study on model systems. *Chem Commun (Camb)*: 1271-3.

Rhee, J.-K., D.-G. Ahn, Y.-G. Kim, and J.-W. Oh. 2005. New Thermophilic and Thermostable Esterase with Sequence Similarity to the Hormone-Sensitive Lipase Family, Cloned from a Metagenomic Library. *Appl. Environ. Microbiol.* 71: 817-825.

Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing

the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* 66: 2541-7.

Rousset, M., L. Casalot, B. J. Rapp-Giles, Z. Dermoun, P. de Philip, J.-P. Bélaich, and J. D. Wall. 1998. New Shuttle Vectors for the Introduction of Cloned DNA in *Desulfovibrio*. *Plasmid* 39: 114.

Salvador, A. C., M. d. C. Santos, and J. A. Saraiva. 2010. Effect of the ionic liquid [bmim]Cl and high pressure on the activity of cellulase. *Green Chemistry* 12: 632.

Sánchez, C. 2009. Lignocellulosic residues: Biodegradation and bioconversion by fungi. *Biotechnology Advances* 27: 185.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74: 5463-5467.

Schipper, C., C. Hornung, P. Bijtenhoorn, M. Quitschau, S. Grond, and W. R. Streit. 2009. Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa*. *Appl Environ Microbiol* 75: 224-33.

Schmeisser, C., H. Steele, and W. R. Streit. 2007. Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol* 75: 955-62.

Schoenfeld, T., M. Patterson, P. M. Richardson, K. E. Wommack, M. Young, and D. Mead. 2008. Assembly of Viral Metagenomes from Yellowstone Hot Springs. *Appl. Environ. Microbiol.* 74: 4164-4174.

Seddon, K. R. 1997. Ionic Liquids for Clean Technology. *Journal of Chemical Technology & Biotechnology* 68: 351.

Shill, K., S. Padmanabhan, Q. Xin, J. M. Prausnitz, D. S. Clark, and H. W. Blanch. 2011. Ionic liquid pretreatment of cellulosic biomass: Enzymatic hydrolysis and ionic liquid recycle. *Biotechnology and Bioengineering* 108: 511.

Sinnott, M. L. 2002. Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews* 90: 1171.

Staskawicz, B., D. Dahlbeck, N. Keen, and C. Napoli. 1987. Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. *glycinea*. *J. Bacteriol.* 169: 5789-5794.

Suen, G., P. J. Weimer, D. M. Stevenson, F. O. Aylward, J. Boyum, J. Deneke, C. Drinkwater, N. N. Ivanova, N. Mikhailova, O. Chertkov, L. A. Goodwin, C. R. Currie, D. Mead, and P. J. Brumm. 2011. The Complete Genome Sequence of *Fibrobacter succinogenes* S85 Reveals a Cellulolytic and Metabolic Specialist. *PLoS ONE* 6: e18814.

Swatloski, R. P., S. K. Spear, J. D. Holbrey, and R. D. Rogers. 2002. Dissolution of cellulose [correction of cellose] with ionic liquids. *J Am Chem Soc* 124: 4974-5.

Tamaru, Y., S. Karita, A. Ibrahim, H. Chan, and R. H. Doi. 2000. A Large Gene Cluster for the *Clostridium cellulovorans* Cellulosome. *J. Bacteriol.* 182: 5906-5910.

Toshiyuki, I., M. Yuichi, A. Yoshikazu, H. Shi-hui, W. Shohei, H. Shuichi, K. Motoi, T. Shigeomi, M. Minoru, and H. Yoshihiko. 2006. Increased Enantioselectivity and Remarkable Acceleration of Lipase-Catalyzed Transesterification by Using an Imidazolium PEG-Alkyl Sulfate Ionic Liquid. *Chemistry - A European Journal* 12: 9228-9237.

Turner, M. B., S. K. Spear, J. G. Huddleston, J. D. Holbrey, and R. D. Rogers. 2003. Ionic liquid salt-induced inactivation and unfolding of cellulase from *Trichoderma reesei*. *Green Chemistry* 5: 443-447.

Unsworth, L. D., J. v. d. Oost, and S. Koutsopoulos. 2007. Hyperthermophilic enzymes--stability, activity and implementation strategies for high temperature applications. *FEBS Journal* 274: 4044-4056.

V. Wintzingerode, F., U. B. Göbel, and E. Stackebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews* 21: 213.

Van de Vyver, S., J. Geboers, P. A. Jacobs, and B. F. Sels. 2011. Recent Advances in the Catalytic Conversion of Cellulose. *ChemCatChem* 3: 82.

van Rantwijk, F., R. Madeira Lau, and R. A. Sheldon. 2003. Biocatalytic transformations in ionic liquids. *Trends in Biotechnology* 21: 131.

Vazana, Y., S. Morais, Y. Barak, R. Lamed, and E. A. Bayer. 2010. Interplay between *Clostridium thermocellum* family 48 and family 9 cellulases in cellulosomal versus noncellulosomal states. *Appl Environ Microbiol* 76: 3236-43.

Vazquez-Figueroa, E., V. Yeh, J. M. Broering, J. F. Chaparro-Riggers, and A. S. Bommarius. 2008. Thermostable variants constructed via the structure-guided consensus method also show increased stability in salts solutions

and homogeneous aqueous-organic media. *Protein Engineering, Design and Selection* 21: 673-680.

Ververis, C., K. Georghiou, N. Christodoulakis, P. Santas, and R. Santas. 2004. Fiber dimensions, lignin and cellulose content of various plant materials and their suitability for paper production. *Industrial Crops and Products* 19: 245.

Voget, S., H. L. Steele, and W. R. Streit. 2006. Characterization of a metagenome-derived halotolerant cellulase. *J Biotechnol* 126: 26-36.

Wang, H.-X., Z.-L. Geng, Y. Zeng, and Y.-M. Shen. 2008. Enriching plant microbiota for a metagenomic library construction. *Environmental Microbiology* 10: 2684-2691.

Warnecke, F., P. Luginbuhl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M. Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. Zhang, M. Hernandez, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz, and J. R. Leadbetter. 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450: 560-5.

Warner, C. D., J. A. Hoy, T. C. Shilling, M. J. Linnen, N. D. Ginder, C. F. Ford, R. B. Honzatko, and P. J. Reilly. 2010. Tertiary Structure and Characterization of a Glycoside Hydrolase Family 44 Endoglucanase from *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 76: 338-346.

Wexler, M., P. L. Bond, D. J. Richardson, and A. W. Johnston. 2005. A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ Microbiol* 7: 1917-26.

Wierenga, R. K. 2001. The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Letters* 492: 193.

Wilson, D. B. 2008. Three microbial strategies for plant cell wall degradation. *Ann N Y Acad Sci* 1125: 289-97.

Wood, B. M., K. K. Chan, T. L. Amyes, J. P. Richard, and J. A. Gerlt. 2009. Mechanism of the Orotidine 5´-Monophosphate Decarboxylase-

Catalyzed Reaction: Effect of Solvent Viscosity on Kinetic Constants. *Biochemistry* 48: 5510.

Wu, Z., and Y. Y. Lee. 1997. Inhibition of the enzymatic hydrolysis of cellulose by ethanol. *Biotechnology Letters* 19: 977.

Yang, L., J. S. Dordick, and S. Garde. 2004. Hydration of enzyme in nonaqueous media is consistent with solvent dependence of its activity. *Biophys J* 87: 812-21.

Yang, Z. 2009. Hofmeister effects: an explanation for the impact of ionic liquids on biocatalysis. *Journal of Biotechnology* 144: 12.

Yang, Z., X.-J. Liu, C. Chen, and P. J. Halling. 2010. Hofmeister effects on activity and stability of alkaline phosphatase. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* 1804: 821.

Yang, Z., Y.-J. Yue, W.-C. Huang, X.-M. Zhuang, Z.-T. Chen, and M. Xing. 2009. Importance of the Ionic Nature of Ionic Liquids in Affecting Enzyme Performance. *J Biochem* 145: 355-364.

Yaoi, K., H. Kondo, N. Noro, M. Suzuki, S. Tsuda, and Y. Mitsuishi. 2004. Tandem Repeat of a Seven-Bladed [beta]-Propeller Domain in Oligoxyloglucan Reducing-End-Specific Cellobiohydrolase. *Structure* 12: 1209.

Yennamalli, R., A. Rader, J. Wolt, and T. Sen. 2011. Thermostability in endoglucanases is fold-specific. *BMC Structural Biology* 11: 10.

Yoon, H.-R. 2007. Two Step Derivatization for the Analyses of Organic, Amino Acids and Glycines on Filter Paper Plasma by GC-MS/SIM. *Archives of Pharmacal Research* 30: 387.

Yoon, M. H., and W. Y. Choi. 2007. Characterization and action patterns of two beta-1,4-glucanases purified from *Cellulomonas uda* CS1-1. *J Microbiol Biotechnol* 17: 1291-9.

Zhang, X.-Z., Z. Zhang, Z. Zhu, N. Sathitsuksanoh, Y. Yang, and Y. H. Zhang. 2010. The noncellulosomal family 48 cellobiohydrolase from *Clostridium phytofermentans* ISDg: heterologous expression, characterization, and processivity. *Applied Microbiology and Biotechnology* 86: 525.

Zhang, Y., and P. S. Cremer. 2006. Interactions between macromolecules and ions: the Hofmeister series. *Current Opinion in Chemical Biology* 10: 658.

Zhao, H. 2005. Effect of ions and other compatible solutes on enzyme activity, and its implication for biocatalysis using ionic liquids. *Journal of Molecular Catalysis B: Enzymatic* 37: 16.

Zhao, H. 2006. Are ionic liquids kosmotropic or chaotropic? An evaluation of available thermodynamic parameters for quantifying the ion kosmotropicity of ionic liquids. *Journal of Chemical Technology & Biotechnology* 81: 877-891.

Zhao, H., G. A. Baker, Z. Song, O. Olubajo, T. Crittle, and D. Peters. 2008. Designing enzyme-compatible ionic liquids that can dissolve carbohydrates. *Green Chemistry* 10: 696.

Zhao, H., G. A. Baker, Z. Song, O. Olubajo, L. Zanders, and S. M. Campbell. 2009. Effect of ionic liquid properties on lipase stabilization under microwave irradiation. *Journal of Molecular Catalysis B: Enzymatic* 57: 149.

Zhao, H., Malhotra, S.V. 2002. Applications of ionic liquids in organic synthesis. *Aldrichim Acta* 35: 75-83.

Zhao, H., S. Xia, and P. Ma. 2005. Use of ionic liquids as ‘green’ solvents for extractions. *Journal of Chemical Technology & Biotechnology* 80: 1089.

Zverlov, V. V., G. V. Velikodvorskaya, W. H. Schwarz, K. Bronnenmeier, J. Kellermann, and W. L. Staudenbauer. 1998. Multidomain structure and cellulosomal localization of the *Clostridium thermocellum* cellobiohydrolase CbhA. *J Bacteriol* 180: 3091-9.

Thank you

… Prof. W. Streit, for ceding this interesting theme to me and for giving me the opportunity and the confidence to work self-dependently in your lab, and furthermore for supporting me throughout the years in your group.

…Prof. B. Saake for being the 2nd assessor of this thesis.

… Uli, for always being ready for interesting and fruitful discussions and support and for reading parts of this work very carefully. Thanks a lot.

… Mirjam for your help regarding ARB

… Christian, Sascha, Patrick and Julia J. for your help concerning smaller and larger problems with my computer

… Julia Juergensen, Anne Himmelberg, Friederike Gründger and Jennifer Chow, „my" diploma students, for your willing and committed contribution to this PhD thesis

… Angela, Martina, Uschi and Christiane for your help regarding lab problems and organisational matters

… Jenny, Dagmar, Steffi, Nine, Gina, Claudia, Patrick, Boris, Uli, Sascha, Mirjam, Arek, Sabine K., Sabine Z., Tina, Katja, Gao, Sandra, Mariita, Christian, Moritz, Andreas, Julia J., Nico, Gesche, Anna, Ines, Christel, Angela and the whole group of the last years for the great atmosphere, for always being ready to help and for the time we shared in and outside the lab. I enjoyed it.

… Julia P. for sharing your experience concerning cellulases and ILs with me

... Markus Mientus and Diana Meske for providing cellulases and cellulolytic clones

… the staff of Hagenbeck's Tierpark i.e. Adriane Prahl, Kai Hoppe and the staff of the zoo in Duisburg for providing the samples

… Sarah Endisch for providing the in ILs dissolved substrates and Merck for providing the ionic liquids

… Hanna, Mama & Dicker. Thank you.

Bestätigung der Korrektheit der englischen Sprache

Hiermit bestätige ich, Jun.-Prof. Dr. Mirjam Perner, dass die englische Sprache in der vorliegenden, von Nele Ilmberger verfassten Dissertation mit dem Titel "Metagenomic cellulases from *Bacteria* linking IL-tolerance, halotolerance and thermostability" korrekt ist.

Hamburg, den 05.07.2012
Ort, Datum

M. Perner
Unterschrift