TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Mikrobiologie

Artificial cellulosome: *in vitro* reconstitution of the cellulosome of *Clostridium thermocellum*

Jan Krauss

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. A. Skerra
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. W. Liebl
	2.	Univ.-Prof. Dr V. Sieber

Die Dissertation wurde am 01.03.2010 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 15.06.2010 angenommen.

Meiner Mutter Elke *1954 †1998

Contents

Abbreviations

ATP	adenosine triphosphate
BglB	β-glucosidase B
BSA	bovine serum abumin
c	concentration
CBM	Carbohydrate Binding Module
CbhA	Cellobiohydrolase A
CE	carboesterase
CelJ	Endoglucanase J
CelR	Endoglucanase R
CelK	Exoglucanase K
CelS	Exoglucanase S
CIAP	calf intestinel alkaline phophatase
CipA	Scaffoldin protein A
CMC	carboxy methyl cellulose
Coh	cohesin
°C	degree Celsius
DNA	deoxyribonucleic acid
DNSA	3,5-dinitrosalycylic acid
DTT	dithiothreitol
Doc	dockerin
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
eq	equimolar
FPLC	Fast Protein Liquid Chromatography
g	gram
GH	glycosyl hydrolase
h	hour
His	histidin
IMAC	immobilized metal ion affinity chromatography
IPTG	isopropyl-beta-D-thiogalactopyranosid
IR	inverted repeats
IS	insertion sequence

kb	kilobases
kDa	kilo Dalton
l	litre
LB	Luria broth
MALDI-TOF	Matrix Assisted Laser Desorption/Ionisation time of flight
MES	3-morpholinoethanesulfonic acid
µg	microgram
µl	microlitre
µM	micromolar
mg	milligram
min	minute
ml	millilitre
mM	millimole
MOPS	3-morpholinopropanesulfonic acid
MW	molecular weight
NHS	*N*-hydroxysulfosuccinimide
Ni	nickel
nm	nanometer
NP	nanoparticle
NTA	N_{α}, N_{α}-bis(carboxymethyl)-L-lysine
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PASC	phosphoric acid swollen cellulose
PCR	polymerase chain reaction
PEG	poly ethylene glycol
PGO	peroxidase-glucose oxidase
RNA	ribonucleic acid
rpm	rounds per minute
RT	room temperature
SDS	sodium dodecyl sulphate
SM	swimmer mutant
SOC	super optimal broth with catabolite repression
t	time
T	temperature

T_m	melting temperature
TLC	Thin layer chromatography
U	unit
UN	united nation
V	volume
v/v	volume/volume
w/v	weight/volume
XghA	Xyloglucanase A
XynA	Xylanase A
XynC	Xylanase C
XynZ	Xylanase Z

Figures

Chapter 1

Introduction

In our current time, the question of future power production in the face of growing industry and rising prosperity, in which the power requirement rises both from industrialist needs and from wealthy households, is ever more prominent. Since oil and deposits are the basic energy sources apart from nuclear energy, we have to rethink our future drastically. The old, non-renewable energy sources are nearing their end and more and more our attention is drawn to other, renewable forms of power production. The use of renewable sources of raw material decreases recycling problems and the rising strain on the atmosphere caused by greenhouse gases (Leschine, 1995).

Plant biomass like lignocellulose consists predominantly of polysaccharides in the form of cellulose and hemicellulose. Lignocellulose is an almost inexhaustible, renewable source of stored energy (Schwarz, 2003). Sufficient knowledge of the natural procedures for dismantling these sugar containing connections forms the basis for the biotechnological applications of special enzyme systems. So far, cellulose degradation has been investigated particularly intensively.

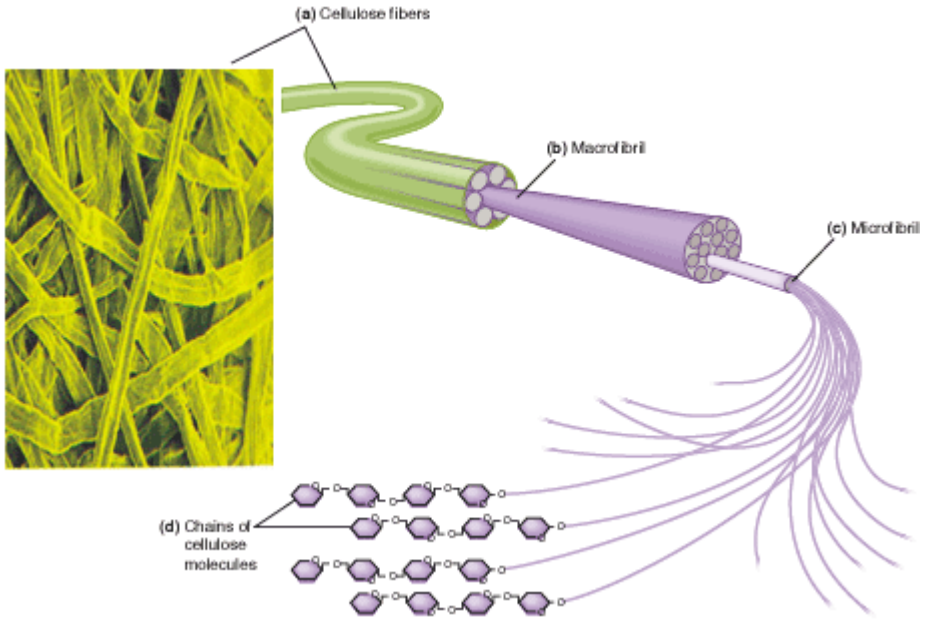
About half of the carbonaceous compounds in terrestrial biomass are cellulose, which is the most prominent single organic compound on earth. The net primary production of biomass was estimated to be 60 Gtons per year of carbon in terrestrial and 53 Gtons per year in marine ecosystems (Cox et al. 2000). Almost all of the biomass produced is mineralized again by enzymes which are provided by microorganisms. Polysaccharide hydrolysis thus is one of the most important enzymatic processes on earth, and cellulose synthesis and hydrolysis is a great part of the carbon cycle.

Thus, there is great incentive to utilize cellulosic biomass as a renewable source of energy via its breakdown to soluble sugars that can then

be converted to liquid fuel (ethanol) or hydrogen (Goldemberg, 2007; Lynd et al., 2008; Xu et al., 2009). The potential quantity of ethanol that could be produced from cellulose is larger than that producible from corn. In contrast to corn-to-ethanol conversion, cellulose conversion involves small or no contribution to the greenhouse effect and has a clearly positive net energy balance (Farrel et al., 2006; Tomme et al., 1995). The main scientific challenge in transforming biomass to liquid fuel is the efficient degradation of plant matter to soluble sugars that can be fermented to liquid biofuel. One promising route involves the exploitation of hydrolytic enzymes of microorganisms that can convert biomass to soluble sugars (Xu et al., 2009; Lynd et al., 2008).

Cellulose is a chemically homogeneous linear polymer of up to 30 000 D-glucose molecules, which are connected by ß-1,4-glycosidic bonds. As each glucose residue is tilted by 180° towards its neighbours, the structural subunit of cellulose is cellobiose (Fig. 1B). The chemical uniformity causes crystallization of the cellulose based on formed H-bonds between the molecules, resulting in tightly packed microfibrils (Fig. 1A). Cellulose is thus a sturdy material ideally suited to insure the structural stability of land plants where it is a main component of the primary cell wall, especially in wood. Due to the high-grade arranged crystalline structure, insoluble cellulose is not completely degradable by individual enzymes (Tomme et al., 1995); a network of hemicellulose, lignin and other polymers in a merged structure is thereby the main cause (Schwarz, 2004). However, the crystals are not perfect but interrupted by amorphous regions. In contrast, soluble cellulose derivatives are easily degraded by a single endo-ß-1,4-glucanase.

A)



B)

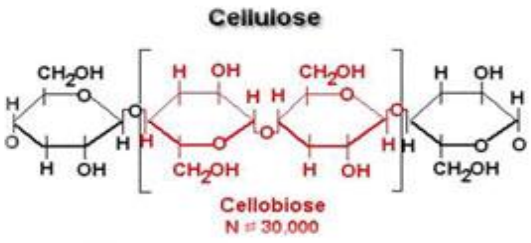


Fig. 1: A) Structure of cellulose.
(a) Cellulose fibers from a ponderosa pine. (b) Macrofibrils compose each fiber. (c) Each macrofibril is composed of bundles of microfibrils. (d) Microfibrils, in turn, are composed of bundles of cellulose chains (nutrition.jbpub.com).

B) Chemical structure of cellulose.
A polymer of D-glucose molecules connected by ß-1,4-glycosidic bonds.

Due to the increasing significance of the hydrolysis of polysaccharides from plant biomass to fermentable sugars, cellulases play an enormous role and are more and more the focus of interest, more so as cellulose is an upcoming substitute for starch as a source of glucose for large scale biotechnology which is based on fermentation. Moreover, the final product of cellulose hydrolysis, glucose, is a universal raw material for a whole pallet of microbial fermentations (Schwarz et al., 2007). So far, the commercially offered cellulase preparations are far from meeting the requirements of an efficient, stable and rapid hydrolysis system (Igelspacher et al., 2006). Today, the advancement of the industrially manufactured fungal cellulases has largely been exhausted. But one can draw upon alternative and efficient cellulase concepts. Cellulases of anaerobic bacteria form a large enzyme complex in contrast to the soluble enzymes of aerobic fungi, which hydrolyse crystalline cellulose with much higher efficiency. These complexes are known as

cellulosomes. One of the best characterized enzyme systems is the cellulosome of the anaerobic thermophilic bacteria *Clostridium thermocellum* (Schwarz, 2001). The cellulosome of *Clostridium thermocellum* represents an extracellular multi-enzyme complex, which is composed of at least 14 different major proteins (Lamed et al., 1983 Bayer et al., 1985). The macromolecular extracellular complex consists of a fibrilar protein anchored to the bacterial cell wall (the scaffoldin unit CipA) and of numerous hydrolytic enzyme subunits. The non-catalytic scaffoldin contains binding sites, the cohesins to which the cellulosomal enzyme units bind with their dockerin modules. Moreover, the scaffoldin unit bears a carbohydrate-binding-module to keep the whole complex in contact with the substrate (Fig. 2A). The cellulosome is considered to be the most efficient of all cellulase systems for cellulose degradation, due to the organization of the enzymes into a complex that "concentrates" them together on given sites of the cellulosic substrate and facilitates stronger synergism among the catalytic units. Indeed, the incorporation of dockerin-bearing cellulases into artificial designer cellulosomes (Fierobe et al., 2001) was shown to induce synergism between cellulases via targeting to the substrate or by the proximity of the cellulases in the complex (Caspi et al., 2009; Fierobe et al., 2002; Fierobe et al., 2005; Mingardon et al., 2007). Sequences of dockerins were included in a BLAST search against the draft genomic sequence. The search identified 72 reading frames containing a dockerin module (Zverlov et al., 2005). This indicates a high number of cellulosomal hydrolases, considering the limited fermentation abilities of *Clostridium thermocellum* (Zverlov et al., 2006).

The aim of this study is the development of an artificial cellulase system based on the structure of the cellulosome from *Clostridium thermocellum*. The technical enzymes for the degradation of plant biomass with fibrous portions (mainly crystalline cellulose) produced at present are compounded from relatively few soluble, non-complexed enzymes. They are produced almost exclusively from fungi such as *Trichoderma reesei*. Cellulase complexes, like those of some anaerobic bacteria e.g. the cellulosome, show a much higher specific activity, especially for crystalline cellulose. This is caused by, among other things, the spatial proximity of differently effective

enzymes at the effect place, which leads to a strong synergism. Thus several enzymes, if they are bound in certain complexes, are more than 15 times more active than if they are present in soluble form (Zverlov et al., 2008). This is not valid for just any composition of cellulases, but requires a function and counterbalancing of the components co-ordinated quantitatively to each other. Such an enzyme complex was compiled artificially here for the disruption of crystalline cellulose, for the modern „white biotechnology" indispensable substrate.

The investigation of the cellulosome was substantially hindered by the impossibility of dividing these complexes into native individual components in order to e.g. set it together afterwards purposefully in a deliberately new combination again. For the decomposition of the complex, extremely strong connections have to be broken; therefore denaturing conditions had to be used. This impaired the folding of the enzymatic components, which were then no longer present in fully active form. These difficulties could be eliminated by the isolation of a mutant of *Clostridium thermocellum*, making the reconstitution of the cellulosome possible.

The basis of this study is this mutant *Clostridium thermocellum* which lacks the macromolecular complex. The mutant exhibits an ORF-interruption of the scaffoldin encoding *cipA*-gene caused by transposon IS 1447 (Zverlov et al., 2008). Therefore cellulosomal hydrolases, which are immobilized on the membrane anchored scaffoldin unit, are secreted in unbound state to the surrounding media. The mutant produced the cellulosomal components in about equal amounts compared to the wild type. Via dockerin-cohesin interaction, these enzymes were connected to hydrolase carriers, so called minicellulosomes. These structures are recombinantly produced fragments of the scaffoldin CipA protein of *Clostridium thermocellum*, miniscaffoldins, composed of different numbers of cohesins (up to nine) and optionally a carbohydrate binding module (CBM). The miniscaffoldins could be immobilized via peptide chemical coupling methods on the surface of modified nanoparticles to create bionic nanocellulosomes (Fig. 2B).

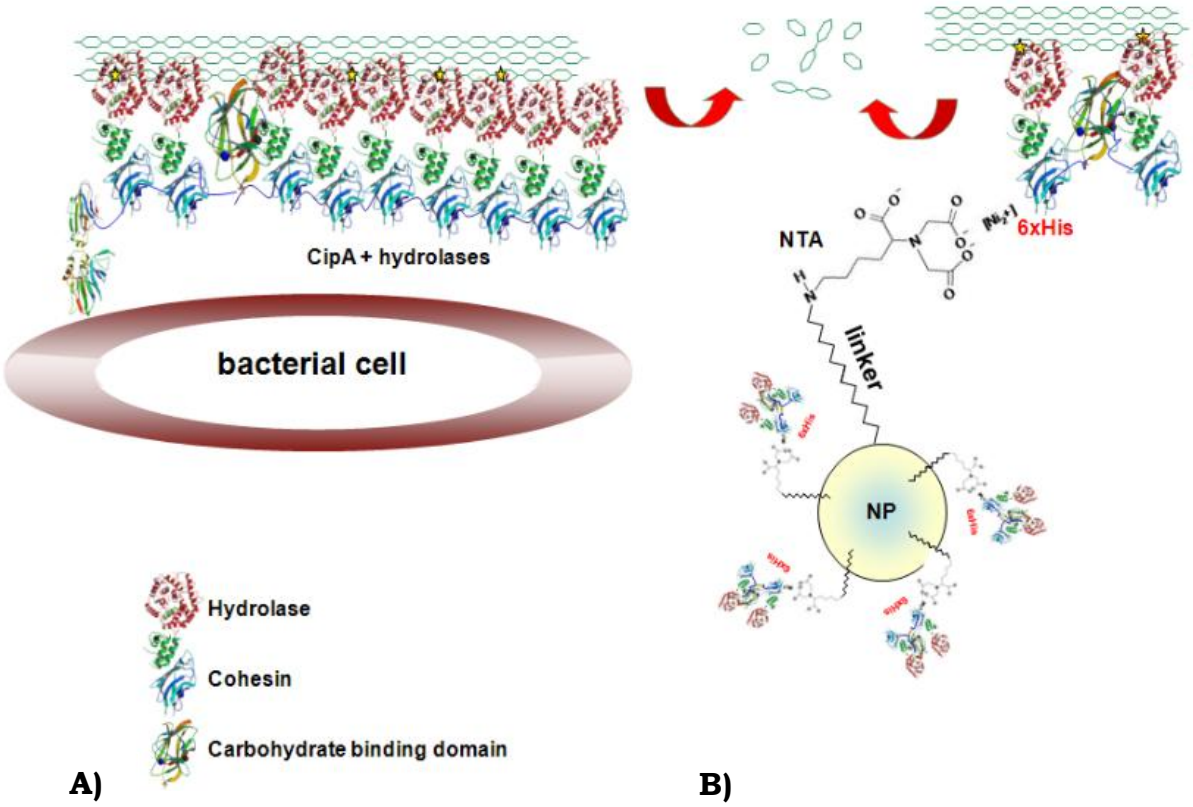


Fig. 2: Comparison of native and artificial cellulosomes. A) Scaffoldin (CipA) is bound via anchoring-protein (SbdA) on bacterial surface; it consists of nine different cohesins with hydrolytic dockerin bearing enzymes and a CBD. **B)** Hydrolases bearing miniscaffoldins are bound on the surface of nanoparticles (NP) via an interconnected linker (protein structures courtesy of Carvalho et al., 2007; Alzari et al. 1996).

The technical challenge is to manufacture high-efficient enzyme complexes for the dismantling of crystalline cellulose by a recombinant scaffoldin protein construct and the soluble components of *C. thermocellum* mutant, which lose the ability for producing CipA protein. The optimal composition necessary for the cellulose dismantling has to be determined and reconstructed by a manufactured mixture of recombinant enzyme components, which can then be used in technical scale.

Furthermore parameters such as the density of enzymes molecules packed onto nanoparticles, pH-, temperature stability of the enzyme complexes and especially hydrolytic activity towards different high crystalline substrates were evaluated. Moreover the composition of the mutant enzyme cocktail was varied by the addition of recombinant cellulosomal cellulases to manufacture high-efficient enzyme complexes for the hydrolysis of

crystalline cellulose. This had the aim of determining the composition necessary for optimal cellulose degradation.

Thus, the developed structures borrow the decoding of innovation of encouraging invention of a naturally superior enzyme complex and its innovative transfer into technology. This bionic system competes with commercially available cellulase preparations in terms of hydrolytic effectiveness, recycling problems and stability. Using the nanoparticle-enzyme complex sugar (glucose), quantities sufficient for industrial procedures can be manufactured. Last but not least raw material extraction for fermentations does not stand in competition with the food and animal feed production.

Chapter 2

Theoretical background

2.1 Mechanism of cellulose hydrolytic enzymes

The well characterized cellulase system of the basidomycete *Trichoderma reesei* consists of endoglucanases, which cleaves glycosidic connections in the amorphous regions of cellulose; cellobiohydrolases (exoglucanases), which split cellobiose gradually from developed non-reducing ends; and β-glucanases, which finally hydrolyze the cellobiose and cello-oligomeres to glucose (Wood and McCrae, 1978) (Fig. 3).

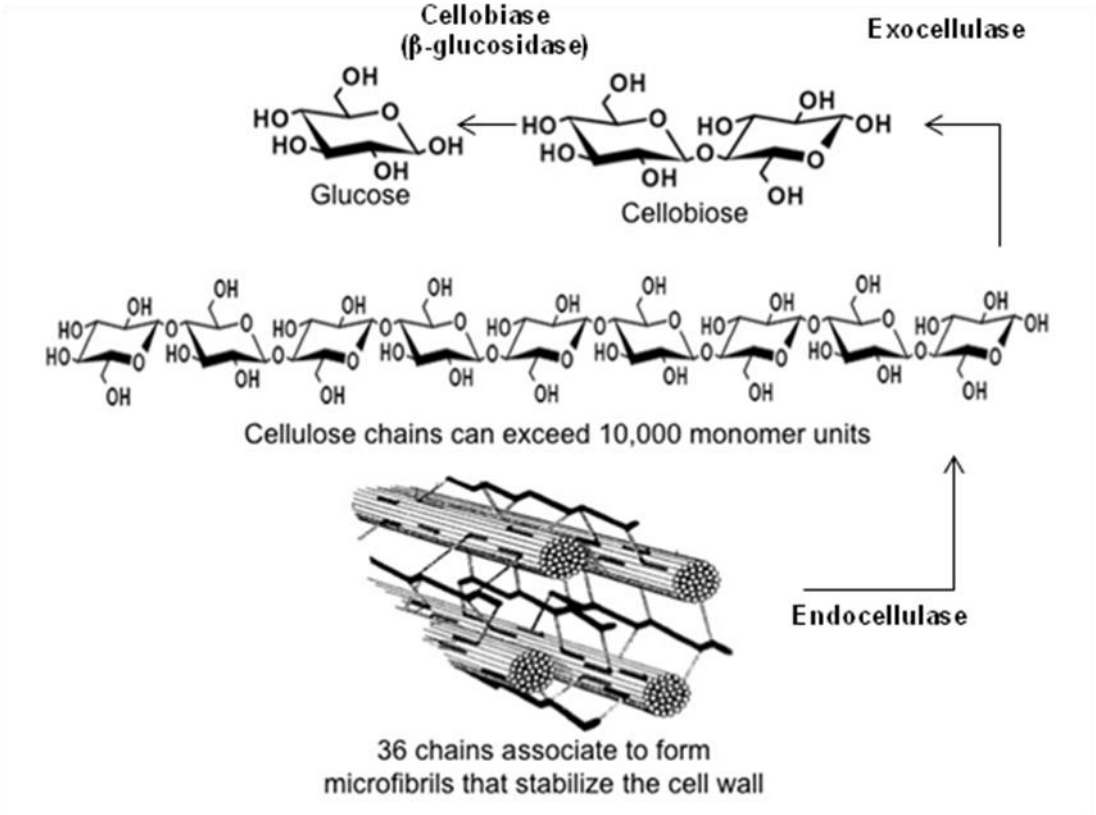


Fig. 3: 3 different types of cellulases and resulting products (www.cchem.berkeley.edu)

The current model of cellulose hydrolysis is:

- an endo-glucanase attaches to the surface of a cellulose crystal (crystalline or amorphous region, edges or planes etc.)
- it pulls one molecule of cellulose from the surface and hydratizes it
- it threads the molecule into the active site pocket, either in an endo-mode or from one of the ends (exo-mode)
- it cuts the glycosidic bond and creates new ends (a non-reducing and a reducing end)
- a processive glycosyl-hydrolase proceeds to the next cut by threading the molecule through the active site pocket, and releases either cellobiose or cellotetraose (cellobiohydrolase or cellotetraohydrolase); a non-processive endo-glucanase will release both ends and move by diffusion to the next site of activity

Processive enzymes can then pick up the ends and degrade a thread of a cellulose molecule on the surface of the crystal. Successive degradation events like the one described above will expose a new surface layer of the crystal for another round of concerted enzyme attack (Fig. 4).

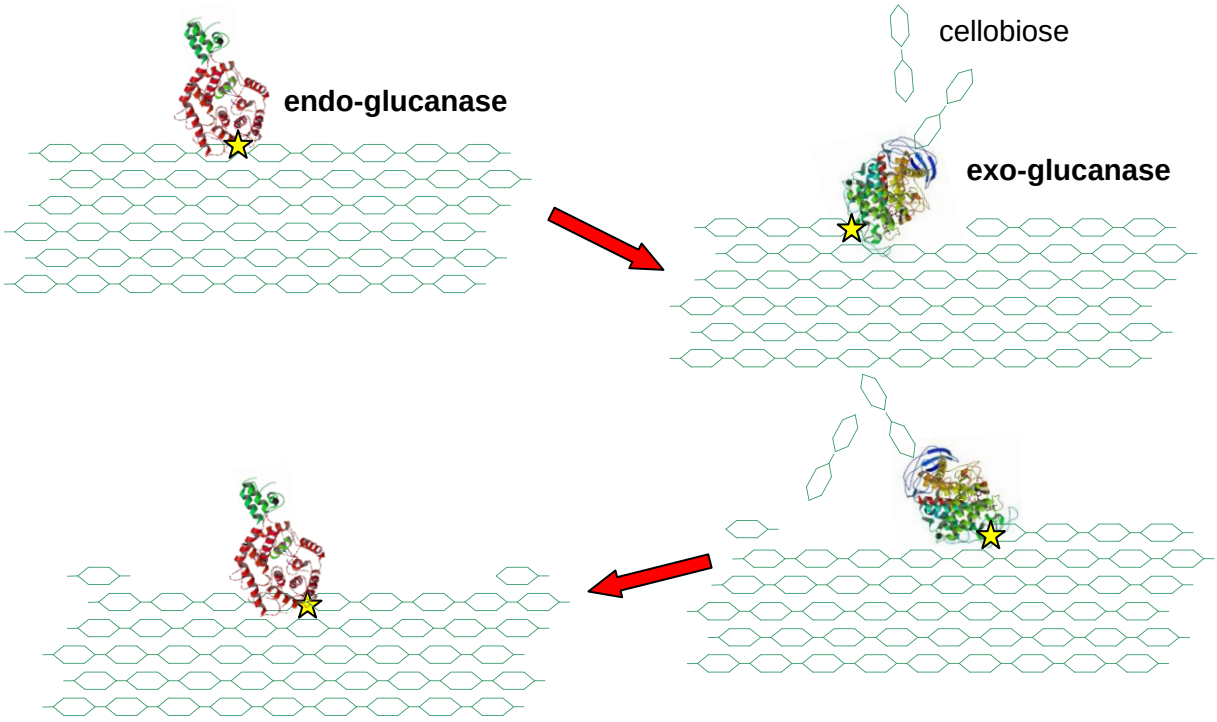


Fig. 4: model of cellulose hydrolysis (protein structures by Alzari et al., 1996; Schubot et al., 2004)

The catalytic modules of the glycosyl-hydrolases were divided from each other by sequence comparison into more than 100 families (www.CAZy.org). The overall 3D-structure and the stereospecificity of hydrolysis are conserved within a family. More than 600 cellulase genes are listed in the CAZy data base so far. Also, enzymes with different substrate spectra were summarized within the same family. Enzymes which convert identical substrates were sometimes found in different families.

The organization of the catalytic centers allows itself to be divided into three forms, despite the differences in the protein. The bag or crater form is found with monosaccharidases, like β-glucosidases and β-galactosidases. The gutter or furrow form is typical for enzymes hydrolysing polysaccharides in endo-mode. The tunnel form, with which the polysaccharide chain is shortened prozessively from the non-reducing end, was found only with cellobiohydrolases, later termed exo-glucanases (Davies and Henrissat, 1995). The hydrolysis of the substrates within the catalytic modules takes place via acid base catalysis, meaning that two amino acid side chains (aspartate or glutamate) are involved. The transient condition of the hydrolysis represents an oxo-carbenium ion, independently of the transient condition of the hydrolysis in which the anomeric C-atom of the substrate molecule changes its conformation (inversion) or remains (retention) (McCarther and Withers, 1994) (Fig. 5).

Inverting enzymes utilize two enzymic residues, typically carboxylate residues, which act as acid and base, respectively (Fig. 5A).

Retaining glycosidases operate through a two step mechanism, with each step resulting in inversion for a net retention of stereochemistry (Fig. 5B). Again, two residues are involved which are usually enzyme-borne carboxylates. One acts as a nucleophil and the other as an acid/base. In the first step the nucleophile attacks the anomeric centre, resulting in the formation of a glycosyl enzyme intermediate, with acidic assistance provided by the acidic carboxylate. In the second step, the now deprotoned acidic carboxylate acts as a base and assists a nucleophilic water to hydrolyze the glycosyl enzyme intermediate, resulting in the hydrolyzed product (Vocadlo et al., 2001).

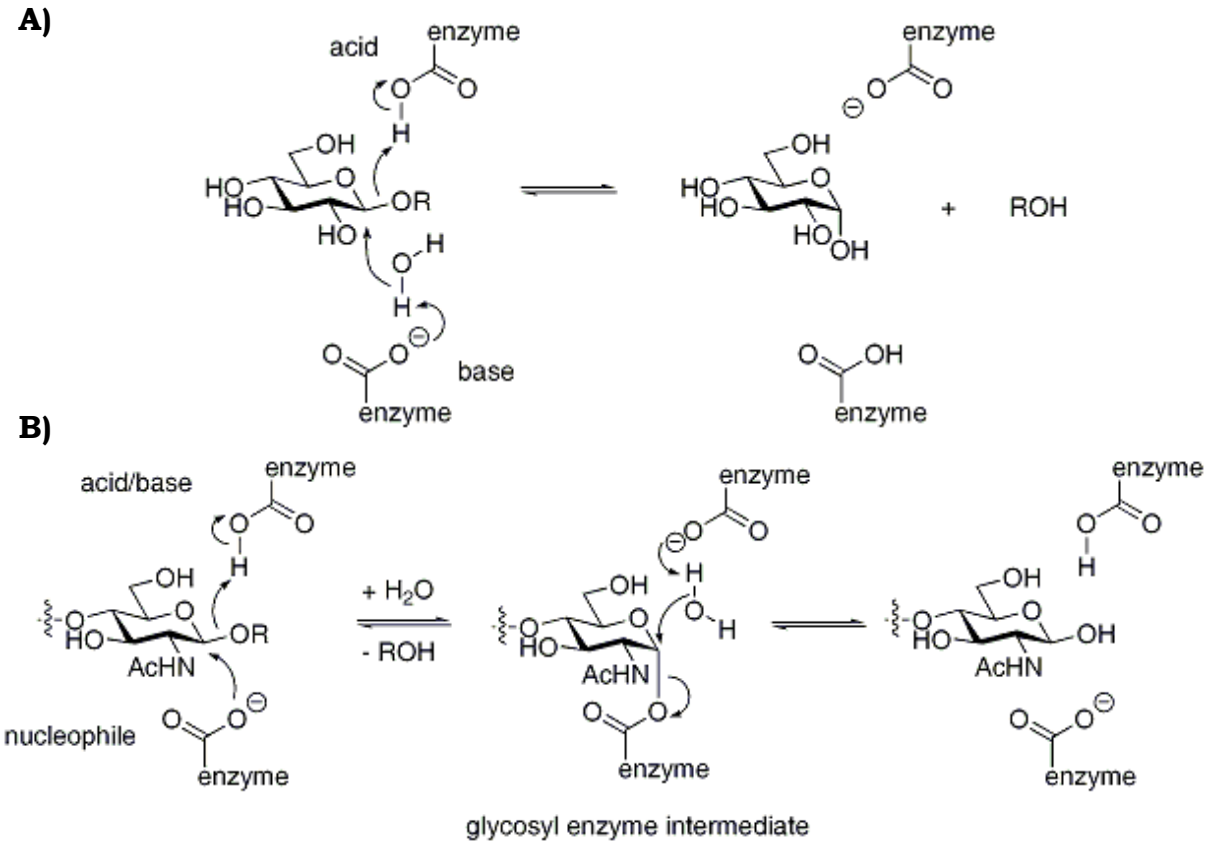


Fig. 5: A) Inverting and B) retaining mechanism for a glycoside hydrolase

An alternative mechanism for hydrolysis occurring with retention of stereochemistry proceeds through a nucleophilic residue that is bound to the substrate rather than being attached to the enzyme. Such mechanisms are common; for instance, for certain N-acetylhexosaminidases, which have an acetamido group capable of neighboring group participation to form an intermediate oxazoline or oxazolinium ion. Again, the mechanism proceeds in two steps (Fig. 6) through individual inversions, leading to a net retention of configuration (en.wikipedia.org/wiki/Glycoside_hydrolase).

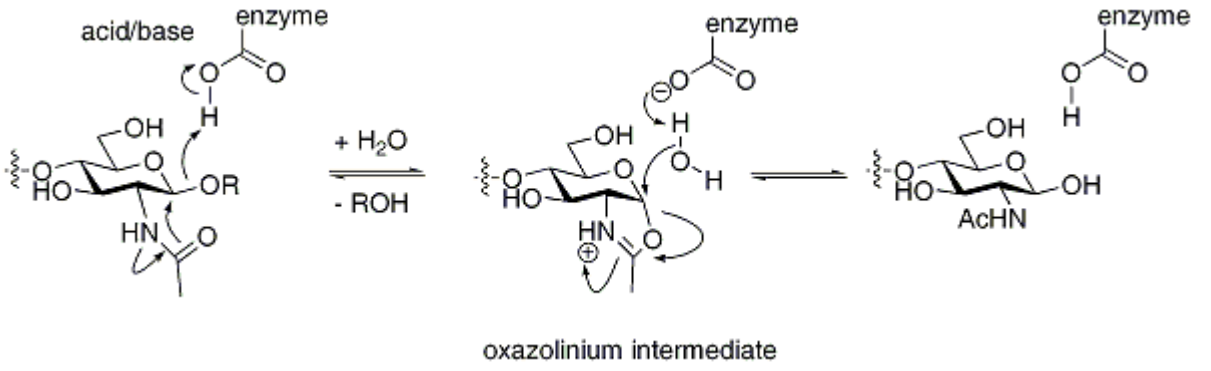


Fig. 6: Alternative retaining mechanism for a glycoside hydrolase

2.2 *Clostridium thermocellum*

(**Domain**) Bacteria; (**Phylum**) Firmicutes; (**Class**) Clostridia; (**Order**) Clostridiales; (**Family**) Clostridiaceae; (**Genus**) Clostridium

C. thermocellum is an anaerobic thermophilic Gram$^+$ bacterium with a rod shaped cell body (Fig. 7). It is the best-characterized thermophilic anaerobic bacterium and is capable of the complete degradation of cellulose (Ljungdahl et al., 1981). Viljoen, Fred and Peterson first described *C. thermocellum* having isolated it from horse manure in 1926 (Viljoen et al., 1928). However, it was 25 years before a pure culture was obtained (Mc Bee, 1948). *C. thermocellum* is widespread in nature; its habitat is decomposing organic material. Investigators have found strains in agricultural waste, sewage digestion sludge, soil, cotton bales, river mud, and hot springs (Ljungdahl et al., 1981; Ng et al., 1977; Bender et al., 1985; Ljungdahl et al., 1983; Wiegel et al., 1985; Zverlov, personal communication).

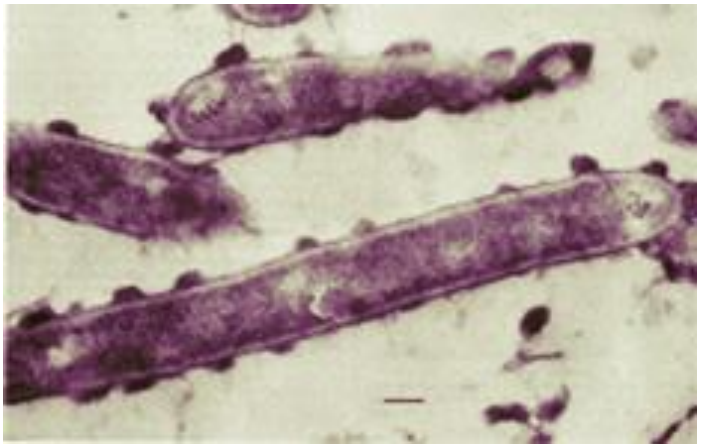


Fig. 7: *Clostridium thermocellum.* (Bayer et al., 1986)

The microorganism grows in complete anaerobiosis and in the thermophilic temperature range. The optimum temperature for growth is 60-64 °C and the optimum pH ranges from 6.1 to 7.5 (Freier et al., 1988). The cells are straight or slightly curved rods, often with tapered ends, and they occur singly or in pairs. The spores are oval, terminal, and swell the cell. Surface colonies are watery, slightly convex, and frequently produce an insoluble yellow pigment. The microorganism grows slowly. When grown on cellulose, the shortest doubling time reported is 7 hours (Wiegel et al., 1986). The

doubling time on cellobiose is 2.5 hours. In a set of batch studies, stationary phase growth was reached after 11 days of fermentation (Maugeri, 1988).

C. thermocellum degrades several forms of cellulose at different rates. Purified or treated cellulose is degraded at a higher rate than the unprocessed polymer in microcrystalline form. Besides cellulose, this microorganism can also degrade hemicellulose, cellobiose, and xylose oligomers. Sugars such as glucose and fructose are metabolized after adaptation of the culture (Freier et al., 1988; Carreira et al., 1983). The enzymes that degrade the monomeric sugars are induced only after a long adaptation time.

The process of cellulose degradation starts with cell attachment to the insoluble substrate (cellulose). The rate of attachment is related to the growth environment. In its natural habitat, *C. thermocellum* is not as easily detached from cellulose as in batch cultivations. In batch systems, inhibitors, such as glucose and its oligomers that may cause detachment, are accumulated in the medium, while in its natural habitat (soil), the inhibitors are readily consumed by other soil microorganisms. Therefore, attachment is induced by cellulose and repressed by large amounts of sugars (glucose and its oligomers). However, other end products, such as ethanol, acetic acid and lactic acid have no effect on attachment (Wiegel et al., 1985).

C. thermocellum remains attached to cellulose and eventually sporulates when the environment is not suitable for rapid growth. Attachment and sporulation are necessary for its survival in open systems, where the favourable conditions for growth (anaero-biosis and thermophilic temperatures) occur sporadically and are interrupted by periods of unfavourable conditions such as aero-biosis or drops in temperature. In open systems, cellulose degradation is linked with growth. After sporulation, the cellulolytic enzymes are degraded by proteases secreted by other microorganisms. The main products of cellulolytic fermentation with *C. thermocellum* are glucose, cellobiose, lactic acid, acetic acid, formic acid, ethanol, CO_2, and H_2 (Freier et al., 1988). The maximum ethanol concentration produced is low (~ 0.05 % w/w or 12 mM). The addition of a second thermophilic anaerobic culture can, however, increase the ethanol

concentration a hundredfold (Carreira, et al., 1983) by metabolizing excess glucose. During growth, some strains preferentially consume cellobiose over glucose (Ng et al., 1982) although glucose is also part of the immediate products of cellobiose metabolism. When *C. thermocellum* is grown with large amounts of cellulose, glucose is accumulated in the medium (Alexander, 1988), while little if any cellobiose is accumulated. Cellobiose can be metabolized by two different pathways: it can be hydrolyzed to two glucose molecules by the action of cellobiase or reacted to glucose and glucose-1-P by cellobiose phosphorylase.

C. thermocellum is best known for its ability to degrade cellulose which is an extremely insoluble compound. However, in order for cellulose degradation to occur, *C. thermocellum* produces many enzymatic proteins that are particularly vital for cellulolysis. Biotechnological research has shown that the cellulose degrading bacteria produce a large, complex cellulase system, known as the cellulosome, which consists of about 20 proteins that are involved in the bacteria's adherence to cellulose, breakdown and regulation of cellulose degradation, and the transport of sugar monomers (http://genome.jgi-psf.org). One of the most important proteins of the cellulosome is the CipA which is a large, non-catalytic ~200 kDa scaffold protein. When the whole genomic sequence of *C. thermocellum* ATCC 27405 became accessible (GenBank accession no. CP000568), 72 cellulosomal genes, including *cipA*, were identified in the genome (Zverlov et al., 2008). As a scaffold protein, CipA is a protein that has multiple specific binding sites which serves to recruit other smaller proteins sharing the same signalling pathway. When brought together by CipA, the proteins can interact to signal and ultimately trigger cellulose degradation. In this case, CipA has nine cohesin modules for protein binding and is mediated by the dockerin modules on the catalytic proteins of the cellulosome. In addition to activation of cellulose degradation CipA also contains cellulose binding factors which are absolutely essential for cellulolysis to even occur in the first place. The cellulose binding factors allow *C. thermocellum* to adhere to the surface of cellulose so that the insoluble substrate can be degraded. Other functions could be speculated to be loosening the crystalline surface by pulling out single substrate molecules (Schwarz, personal communication). Aside from

the scaffold protein in the cellulosome, *C. thermocellum* also produces glycosyl hydrolases which function as the catalytic subunits that engage in the actual cleavage of cellulose (Newcomb et al., 2007).

2.3 The cellulosome of *Clostridium thermocellum*

The cellulosome was first described by Lamed and co-workers. It is a large multi-component cellulolytic complex. Electron-microscopy studies have shown that extracellular protuberances are associated with some cellulose degrading bacteria (Fig. 8) and contain material reacting with antibodies raised against cellulosomal components. These protuberances were predicted to form fibrous contact corridors after adhesion of the cells to cellulose (Bayer et al., 1986). Ultrastructural evidence showed the multi-subunit composition and dimensions of the cellulosome and their organization on the cell surface in the form of poly-cellulosomal organelles (Bayer et al., 1998). Such architecture apparently renders the complex resistant to the most unfriendly conditions.

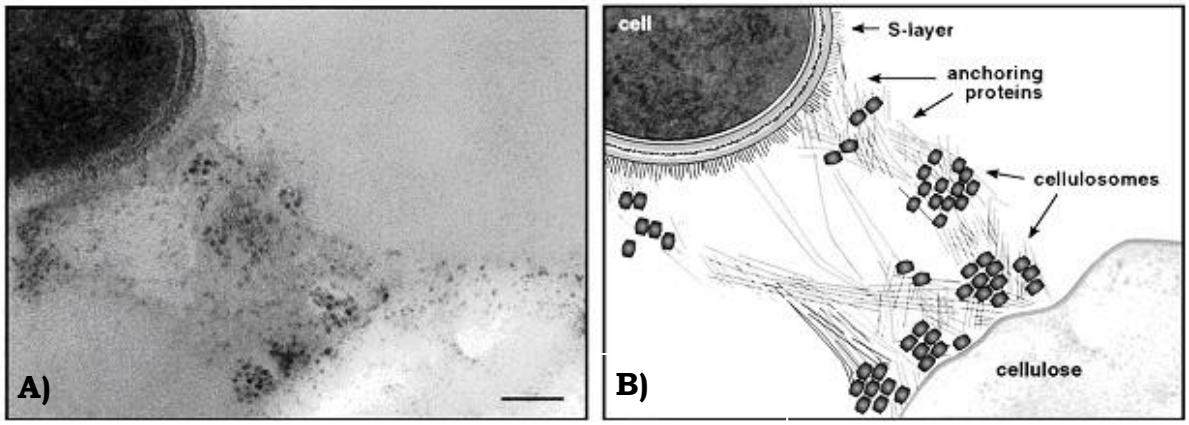


Fig. 8: Ultrastructure of *Clostridium thermocellum* cell surface. A) A high-resolution magnification of a protracted, antibody-labelled poly-cellulosomal protuberance. The cellulosome-specific label is mainly associated with the cellulose surface and connected to the cell via extended fibrous material. **B)** Schematic interpretation of the cellulose-bound cell surface (figures: Bayer et al., 1998). Bar = 100nm.

This structure was reported to have a mass of $2.1*10^6$ Daltons and initially 14 different polypeptides were found (Lamed et al., 1983). Its multi-functionality represents the efficiency of the evolutionary processes that provided clostridia and other bacteria with a mechanism that enables cells to obtain energy from the two most abundant but intrinsically intractable substrates (Felix et al., 1993), cellulose and hemicellulose.

1983 Lamed et al. defined the cellulosome as a discrete, cellulose binding, multi-enzyme complex for the degradation of cellulosic substrates. This definition was based on *Clostridium thermocellum* structure, one of the first cellulosomes to be studied in detail. The cellulosome enables the connection

of the catalytic components and the microorganism to the substrate and the synergistic cooperation of different cellulolytic enzymes in spatial coordination to each other (Béguin et al., 1996). Cellulosomal enzyme subunits are not different from related free cellulases; both free and cellulosomal enzymes contain common types of catalytic modules from the same collection of glycosyl hydrolase families. The major difference between the two types of enzymes is that all cellulosomal enzymes contain a module which mediates the enzymes's integration into the cellulosome complex, whereas noncellulosomal enzymes apparently lack such a domain (Bayer et al., 1994; Béguin and Lemaire, 1996).

Molecular-biological methods opened the view of the special role of functional protein modules, which are responsible for the structural characteristics of the cellulosome. The modules exhibit thereby highly specific protein-protein and protein-substrate interactions. A majority of *C. thermocellum´s* polysaccharide splitting enzymes possess a duplicated motive for sequence over 22 amino acids. Within this repeated sequence appears a 12 residue segment that exhibits sequence similarity with the EF-hand motif of calcium-binding proteins, e.g., calmodulin and troponin C (Chauvaux et al., 1990). Proteins carrying this element interact with the cellulosomal scaffoldin unit (Tokatlidis et al., 1991). The duplicated element is designated according to its function as dockerin module (type I). The determination of the primary structure of the non-catalytic scaffoldin unit showed that it consists of a 196 kDa protein of nine repetitive motives with a length of 170 AS. Each one can bind a protein with a dockerin, these are called cohesin (type I) modules.

Haimovitz et al. (2008) developed a protein micro-array system for evaluating the binding specificity of test dockerins for a library of cohesin modules, enabling a more comprehensive investigation of cross- and intra-species cohesin–dockerin interactions. The intra-species fidelity of *Clostridium thermocellum* was demonstrated, with the test dockerin (Ct-Doc48S) interacting with nine type I cohesins from the CipA scaffoldin of *C. thermocellum*. The dockerin did not interact with any type-II cohesins or type-I cohesins from other species. Weak cross-species binding was observed rarely, such as the type II dockerin and cohesin of *Clostridium josui* and

Clostridium cellulolyticum (Barak et al., 2005). The type II dockerin is found at the C-terminal region of CipA. It shows conformational flexibility (Adams et al., 2005) and interacts specifically with the dockerin-bearing membrane anchoring protein SdbA to display the complex on the cellulolytic organism's membrane. Thus adaptor and anchoring scaffoldins ensure the juxtaposition of the complex between the bacterium and plant cell wall.

Calcium is required for the folding and stabilisation of the cohesin–dockerin interaction. Type I dockerins possess two calcium binding segments which are required for cohesin recognition and species specificity (Schaeffer et al., 2002). The solution structure determined by NMR indicates that dockerins undergo substantial conformational change upon binding to cohesins. The crystal structure of the complex showed that cohesin recognition was predominately through helix-3 of the dockerin, whereas the almost perfect duplication of the 22-residue sequence in helix-1 could result in reverse binding of the dockerin (Carvalho et al., 2007). The cohesin-dockerin complex formation is one of the strongest protein-protein interaction known so far; the components interact strongly with an affinity constant of the order of 10^9 M^{-1} (Fierobe et al., 1999).

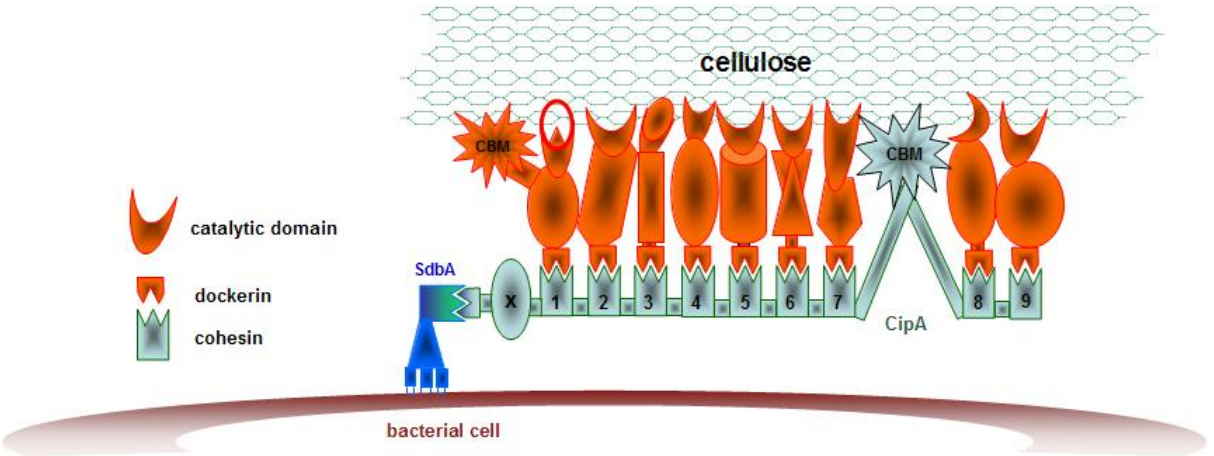


Fig. 9: Conceptual diagram of cellulosome architecture. The cellulosome co-localises catalytic modules for degradation of plant cell walls. The scaffoldin backbone of the cellulosome is flexible, and can conform to heterogeneous cellulosic geometries. The primary scaffoldin includes several cohesin modules as well as a cellulose carbohydrate binding module (CBM). Adaptor scaffoldins contain both cohesin and dockerin modules. The anchoring scaffoldin includes a SLH (S-layer homology, SdbA) module. The cohesin modules are numbered 1 to 9. The cellulosomal components are not drawn to scale.

The high affinity, 'lock-and-key' interaction between cohesins and dockerins specifies cellulosome architecture, and dockerin bearing cellulases are

assembled onto a scaffoldin backbone (CipA protein) via high-affinity, calcium-dependent interactions. Enzymatic subunits of the complex are 'tagged' by a dockerin module, which serves to locate catalytic subunits at specific locations on the scaffoldin (Fig. 9).

Its components display significant synergism since the individual proteins have low activity against cellulose, presumably because they lack the CBMs necessary to bring catalytic sites into close proximity with the insoluble substrate (Blumer-Schuette et al., 2008). Thus, the advantage of the enzyme complex is the presence of different enzymes in a high local concentration and in spatial proximity to the substrate, whereby fewer enzymes have to be produced and energy can be saved for the organism.

The primary scaffoldin contains a module for binding the complex to cellulose, the so-called cellulose carbohydrate binding module (CBM) (Nordon et al., 2008). It is defined as contiguous amino acid sequence within a carbohydrate-active enzyme (or within a scaffoldin protein) with a discreet fold having carbohydrate-binding activity. CBM have evolved, mostly by soil bacteria and fungi, for the degradation of plant biomass. CBMs were previously classified as cellulose-binding domains (CBDs) based on the initial discovery of several modules that bound cellulose. However, additional modules in carbohydrate-active enzymes are continually being found that bind carbohydrates other than cellulose yet otherwise meet the CBM criteria, hence the need to reclassify these polypeptides using more inclusive terminology (Tomme et al., 1988; Gilkes et al., 1988). Thus they were divided into different families. For example, cellulases with a type B CBM interact with single polysaccharide chains (Boraston et al., 2001). Type A CBM (e.g., Family I and III) promotes hydrolysis at different sites on crystalline cellulose (Carrard et al., 2000). The structure protein of the cellulosomes from *C. cellulovorans, C. thermocellum and C. cellulolyticum* has a single type A III CBM (Tomme et al., 1998).

The Family-IIIa CBM of the scaffoldin is a "conventional" CBM, which exhibits recognition of and strong binding to crystalline cellulose (Morag et al., 1995). As a consequence of its action, this CBM mediates the primary recognition and binding of the scaffoldin subunit (along with its attached cellulosomal enzymes) to the cellulosic substrate. When the cellulosome is

implanted in the cell surface, the scaffoldin CBM thus mediates the binding of the entire cell to the insoluble substrate (Bayer et al., 1996).

Several cellulosomal enzymes also bear CBMs as part of their structure, although the presence of a CBM seems not to be a definitive cellulosomal characteristic, and its function is not necessarily that of a targeting agent. For this purpose, the cellulosomal enzymes depend collectively on a special CBM, borne by a separate scaffolding subunit (Bayer et al., 1998). Aerobic cellulolytic species such as *Trichoderma reesei*, and anaerobic species as *C. thermocellum* also secretes monomeric enzymes with a CBM (Warren, 1996; Harhangi et al., 2003) which mediates the binding of free hydrolase to the substrate. The essential function of such a CBM was shown for the cellobiohydrolase CBHI from *T. reesei*, for which a detailed 3-dimensional model was constructed (Lee and Brown 1997). The catalytic module without the CBM (the core enzyme) has a very limited overall-action on cellulose.

Until 2005, 28 hydrolytically active cellulosomal components in *C. thermocellum* have been identified solely by screening genomic libraries for a limited selection of hydrolytic activities (Schwarz, 2001). However libraries have not been screened for pectinases, xyloglucanases, esterases, or glycosidases, and their genes have not yet been isolated (Zverlov et al., 2005). Meanwhile, more than 70 cellulosomal components of the *C. thermocellum* cellulosome are known. The most important and most frequently occurring cellulosomal hydrolases of *C. thermocellum* are listed in table 1.

Enzyme	Function	Molecular mass [kDa]	Modular structure
CbhA	Cellobiohydrolase	138	CBD4-Ig-GH9-X-X-CBD3-DS1
CelA	Endoglucanase	53	GH8-DS1
CelB	Endoglucanase	64	GH5-DS1
CelD	Endoglucanase	72	Ig-GH9-DS1
CelE	Endoglucanase	90	GH5-DS1-CE2
CelF	Endoglucanase	82	GH9-CBD3-DS1
CelG	Endoglucanase	63	GH5-DS1
CelH	Endoglucanase	102	GH26-GH5-CBD11-DS1
CelJ	Cellulase	178	X-Ig-GH9-GH44-DS-X
CelK	Cellobiohydrolase	101	CBD4-Ig-GH9-DS1
CelN	Endoglucanase	82	GH9-CBD3-DS1
CelO	Cellobiohydrolase	75	CBD3-PT-GH5-DS1
CelP	Endoglucanase	58	GH9-DS1
CelQ	Endoglucanase	80	GH9-CBD3-DS1
CelS	Exoglucanase	83	GH48-DS1
CelT	Endoglucanase	69	GH9-DS1
ChiA	Chitinase	55	GH18-DS1
LicB	Lichenase	38	GH16-DS1
ManA	Mannanase	67	CBD4-GH26-PT-DS1
XynA (XynU)	Xylanase	74	GH11-CBD4-DS1-CE4
XynB (XynV)	Xylanase	50	GH11-CBD6-DS1
XynC	Xylanase	70	X-GH10-DS1
XynD	Xylanase	70	CBD22-GH10-DS1
XynY	Xylanase/feruloyl esterase	120	CBD22-GH10-CBD22-DS1-CE1
XynZ	Xylanase/feruloyl esterase	92	CE1-CBD6-DS1-GH10

Tab. 1: Most prominent cellulosomal dockerin type I containing hydrolases. The modular structures of cellulosomal subunits are indicated by abbreviations: CBD, cellulose-binding module; CE, carbohydrate esterase family; DS1, dockerin module type I; GH, glycosyl hydrolase; Ig, immunoglobulin-like module; PT, proline-rich linker;; X, unknown module containing a hydrophilic module (Zverlov et al., 2005).

The mechanism by which catalytic subunits are incorporated into the cellulosome has not been established yet. Depending on the kind of substrate, different types of cellulosomes are produced (Blumer-Schuette et al., 2008). For example GH9 (glycosyl hydrolases family 9) cellulases like Cel9R, Cel9J and Cel9K were the most abundant group of enzymes per CipA when cells were grown on cellulose, while hemicellulases were the most abundant group on cellobiose adapted culture (Gold et al., 2007).

2.4 *Clostridium thermocellum* mutant SM901

Zverlov et al. (2008) isolated *C. thermocellum* mutants that had lost the ability to adhere to crystalline cellulose. Six of them showed diminished ability to depolymerize crystalline cellulose. Size exclusion chromatography of the proteins from the culture supernatant revealed the loss of the supramolecular enzyme complex, the cellulosome. Sequence analysis manifested an ORF-interruption of the encoding region of the scaffoldin encoding *cipA* gene caused by an insertion of transposon IS 1447. The IS sequences were identical and consisted of a transposase gene and the inverted repeats IRR and IRS. The insertion resulted in an obviously non-specific duplication of 3 base pairs within the target sequence. This lack of specificity allows transposition without the need of a defined target DNA sequence. Eighteen copies of IS*1447* were identified in the genomic sequence of *C. thermocellum* ATCC 27405 (Zverlov et al., 2008).

The transposon was inserted into the *cipA* reading frame in four different locations, one in each mutant: cohesin module 1, two different positions in the carbohydrate binding module, and cohesin module 3 (Fig. 10).

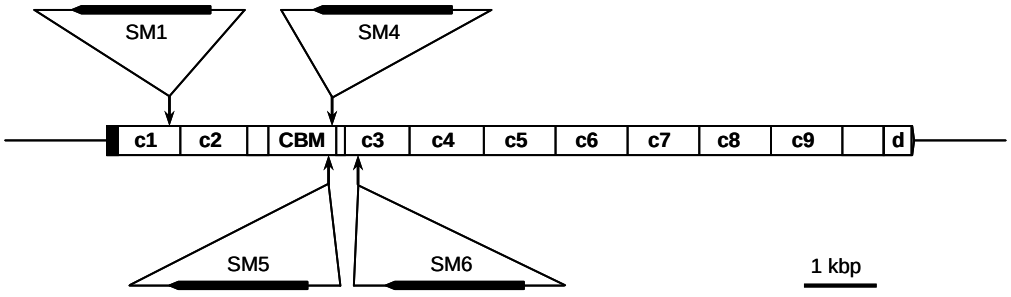


Fig. 10: Structure of the *cipA* gene and positions of IS*1447* insertions (mutant designations are indicated). c1: cohesin module 1; CBM: carbohydrate binding module; d: dockerin module (Zverlov et al., 2008).

Compared to the wild type, the mutant culture supernatant with a completely defective CipA protein showed equal specific hydrolytic activity against soluble β-glucan but a 15-fold reduction in specific activity with crystalline cellulose (Zverlov et al., 2008).

For this study we selected the mutant SM901, which exhibits the insert in the ORF-region of cohesin 1, so the complete scaffoldin CipA protein is not expressed and the cellulosomal enzymes are released in free unbound state

in surrounding media. The mutant which does produce the complete set of cellulosomal enzymes provides a tool to reconstitute the original efficiency of a cellulase complex by constructing artificial cellulosomes.

2.5 Biofuels of 1st and 2nd generation

The control of the global biosphere by human beings is unprecedented in the history of our planet, and our impact is such that substantive changes in ecosystems, and the global environment as a whole, are now becoming apparent. Our activity drives the steady increase in global temperature observed in recent decades. The realization of the adverse effects of greenhouse gas emissions on the environment, together with declining petroleum reserves, has ensured that the quest for sustainable and environmentally benign sources of energy for our industrial economies and consumer societies has become urgent in recent years. Consequently, there is renewed interest in the production and use of fuels derived from plants. First-generation biofuels are biofuels made from sugar, starch, vegetable oil, or animal fats using conventional technology (UN biofuels report). The basic feedstocks for the production of first generation biofuels are often seeds or grains such as wheat, which yields starch that is fermented into bioethanol, or sunflower seeds, which are pressed to yield vegetable oil which can be used in biodiesel. These feedstocks could instead enter the animal or human food chain, and as the global population has risen and continues to rise, their use in producing biofuels has been criticised for diverting food away from the human food chain, leading to food shortages and price increases.

Supporters of biofuels claim that a more viable solution is to raise political and industrial support for the implementation of second-generation biofuels made from non food crops. These include waste biomass, the stalks of wheat, corn, wood, and special energy or biomass crops (e.g. Miscanthus). Second generation biofuels, produced from cheap and abundant plant biomass, are seen as the most attractive solution to this problem, but a number of technical hurdles must be overcome before their potential is realized (Gomez et al., 2008). Second generation biofuels use biological hydrolysis of cellulose and additional fermentation, or chemical biomass to liquid technology (Inderwildi and King, 2009), including cellulosic biofuels

from non food crops (Sommerville, 2008). Many second generation biofuels are under development, such as biohydrogen, biomethanol, DMF, Bio-DME, Fischer-Tropsch diesel, biohydrogen diesel, mixed alcohols and wood diesel. Cellulosic ethanol production uses non food crops or inedible waste products and does not divert food away from the animal or human food chain. Lignocellulose is the "woody" structural material of plants. Producing ethanol from cellulose is a difficult technical problem to be solved. In nature, ruminant livestock ingest grass and use slow enzymatic digestive processes in symbiosis with bacteria to break it down into glucose. The steps for production of second generation biofuels include the pre-treatment of biomass, saccharification and the fermentation of sugars to form fuels such as ethanol (Fig. 11).

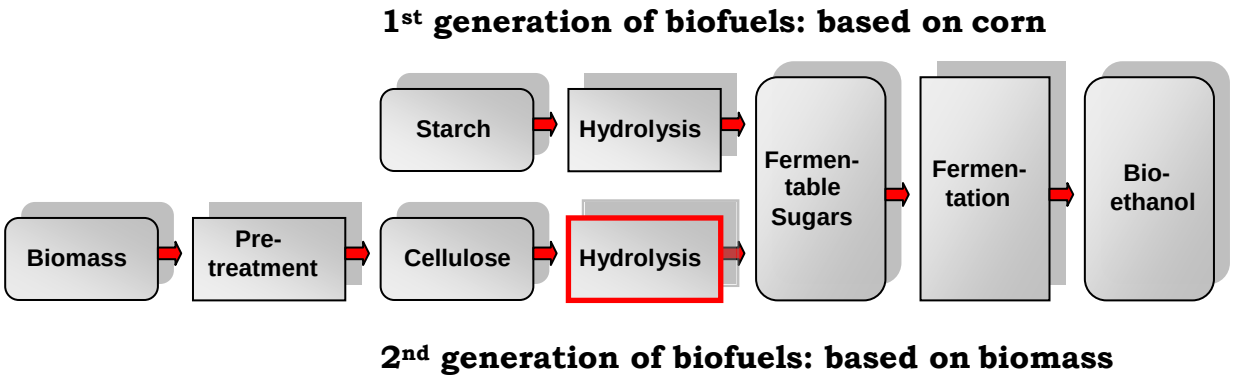


Fig. 11: Overview of bioethanol production according 1st and 2nd generation of biofuels. The process corresponding to this work is marked red.

The commercially produced cellulases used for saccharification are a cocktail of several enzymes that together convert cellulose to simple sugars, the most commonly used of which are sourced from *Trichoderma reesei*. The two key processes in this cocktail are endoglucanases, which attack glucan chains along the microfibril surface leading to reduced polymer lengths, and exo-glucanases, which attack the polymers from their ends (Gomez et al., 2008). Enzymes need to be applied in large quantities and cannot be re-obtained after hydrolysis. Fresh enzymes must be added after every process cycle.

In this study, an enzyme system was developed for the effective degradation of high crystalline cellulose. Hydrolytic enzymes for the dismantling of biomass were immobilized on surface modified nanoparticles. Used enzymes could thus be recycled by an external magnetic field due to the

superparamagnetic behaviour of the particles. Furthermore, commercial products are only available in a free unbound state. Compared to these commercial cellulases, the system proposed in this study consists of closely neighbouring enzymes immobilized on nanobeads. Hence the degradation rate towards crystalline cellulose could be enhanced due to synergistic behaviour.

Chapter 3

Material and methods

3.1 Strains and plasmids

In the following tables the organisms and plasmids used in this study, their genotypical characteristics and their origins are listed.

Strain	Genotype	Reference
Escherichia coli XL1-blue	*supE*44, *hsdR*17, *endA*1, *recA*1, *gyrA*96, *thi*1, *relA*1, *lac*-, F´[*proAB*+, *lac*I^q, *lacZ*ΔM15, Tn*10* (Tet^R)]	Bullock et al., 1987
Escherichia coli DH5α	*supE*44, ΔlacU169, (Φ80 lacZΔM15), *hsdR*17, *endA*1, *recA*1, *gyrA*96, *thi*1, *relA*1	Hanahan, 1983
Escherichia coli M15	F-, *lac*-, *ara*-, *mtl*-, pREP4 (Km^R)	Qiagen
Escherichia coli Top10F´	F´,{*lac*I^q Tn*10* (Tet^R)}, *mcrA*, Δ(*mrr-hsd*RMS-*mcr*BC), Φ80*lac*ZΔM15, Δ*lac*X74, *rec*A1, *ara*D139Δ(*ara-leu*)7697, *gal*U, *gal*K, *rps*L, (Str^R), *end*A1, *nup*G	Invitrogen
Escherichia coli Rosetta gami B (DE3)pLysS	F-, *ompT,hsdS*$_B$ (r_B^- m_B^-), *gal*, *dcm*, *lacY1aphC*, *gor*522::Tn*10*(Tc^R), *trxB*::kan, (DE3)pLysSRARE(Cm^R)	Novagen
Clostridium thermocellum	wildtype	DSM 1237
Clostridium thermocellum SM901	*cipA*::IS1447	Zverlov et al., 2008
Thermotoga neapolitana	wildtype	DSM 4359

Table 2: used microorganisms

Plasmid	Abilities	Resistance	Reference
pQE-30	T5 expression vector N-terminal Histag	Ap^R	Qiagen
pQE-31	T5 expression vector N-terminal Histag	Ap^R	Qiagen
pQE-32	T5 expression vector N-terminal Histag	Ap^R	Qiagen
pET-101/D-TOPO	T7 expression vector C-terminal Histag	Ap^R	Invitrogen
pCoh1	cohesin1 in pQ-E32	Ap^R	this work
pCoh6	cohesin6 in pQE-30	Ap^R	this work
pCoh2-CBM	cohesin2-CBM in pQ-E31	Ap^R	this work
pCoh1-2	cohesin1-2 in pQE-32	Ap^R	this work
pCBM-Coh3-4	CBM-cohesin3-4 in pQE-31	Ap^R	this work
pCoh2-CBM-3	cohesin2-CBM-3 in PQE-31	Ap^R	this work
pCoh1-2-CBM-3	cohesin1-2-CBM-3 in pQE-32	Ap^R	this work
pCoh1-2-CBM-3-4	cohesin1-2-CBM-3-4 in pQE-32	Ap^R	this work
pCipA	scaffoldin CipA in pET-101/D-TOPO	Ap^R	this work
pCbh9A	cellobiohydrolase A in pQE31	Ap^R	this work
pBglB (TN)-Doc	β-glucosidase (*T. neapolitana)* + dockerin in pQE-32	Ap^R	this work
pCel9J	cellulase J in pQE-32	Ap^R	this work
pCel9K	cellobiohydrolase K in pQE-32	Ap^R	this work
pCel9R	endogluconase R in pQE-31	Ap^R	this work
pCel48S	exogluconase S in pQE-32	Ap^R	this work
pG-KJE8	chaperone coexpression vector	Cm^R	Takara Bio
pXynA	Xylanase A in pQE-30	Ap^R	this work
pXynC	Xylanase C in pQE-32	Ap^R	this work
pXynZ	Xylanase Z in pQE-32	Ap^R	this work
pXghA	Xyloglucanase A in pQE-32	Ap^R	this work

Tab. 2: used plasmid vectors

Chemicals were purchased from Sigma-Alldrich, St. Louis, USA, if not mentioned otherwise.

3.2 Preparation of inoculum

All culture media were autoclaved 20 min with 121 °C. Temperature sensitive additives were sterile filtered and added before use.

3.2.1 Culture medium for *Clostridium thermocellum* SM901and wildtype

GS-2-medium (Johnson et al., 1981)	
Cellobiose	5 g
KH_2PO_4	1.5 g
K_2HPO_4	2.9 g
Sodiumcitrate · 2 H_2O	3.0 g
Urea	2.1 g
MOPS	10.0 g
Yeast extract	6.0 g
Resazurin	2 mg
L-Cysteine	1.0 g
H_2O_{dest}	ad 990 ml
	pH 7.2 with 5M NaOH
10x salt solution:	
$MgCl_2$ · 6 H_2O	10 g
$CaCl_2$ · 2 H_2O	1.5 g
$FeSO_4$ · 7 H_2O	10.25 mg
H_2O_{dest}	ad 100 ml

Oxygen had to be removed from the growth media for growing anaerobic. Under nitrogen atmosphere 47.5 ml of GS2 medium in 100 ml serum bottles were sealed by rubber stoppers and autoclaved. By the addition of cysteine as reducing agent the remaining traces of oxygen in the medium were reduced to water. Resazurin was added as redox indicator for reducing conditions. It changes its colour after contact with oxygen from colourless to red (point of transaction: -42 mV). To avoid salt precipitation caused by

autoclaving, the salt solution was added after sterilisation through a sterile filter. Cellobiose was used as substrate. Therefore 2.5 ml of a 10 % solution were added to the media through a sterile filter to get a final concentration of 0.5 % cellobiose.

For inoculation 1 ml of a pre-culture of *Clostridium thermocellum* was injected with a 1 ml syringe and needle in preheated GS-2-medium (Fig. 12). The flasks were incubated for several time periods at 60 °C. The pre-culture was prepared by heating frozen sporulated cultures in GS-2-medium to 80 °C for 20 min to induce spore activation.

Fig. 12: GS-2-medium in 100 ml serum bottles. Right) Pure cellobiose containing GS-2-medium L**eft)** GS-2-medium inoculated with SM901 *C. thermocellum* mutant cultivated after 2 days at 60 °C

3.2.2 Culture medium for *Escherichia coli*

LB-medium (Luria, 1960)	
Trypton	10g
Yeast extract	5g
NaCl	10g
H_2O_{dest}	ad 1l
	pH 7.4

Optionally agar-agar was added to a concentration of 1.8 % (w/v) to the medium. The inhibitors (antibiotics) and medium additives (table 3) were sterile-filtered and added after cooling of the media down to 50 °C (fixed media) or before inoculation (liquid media).

Medium additives	Stock solution	Working concentration
Ampicillin	100 mg/ml in H_2O_{dest}	100 µg/ml
Kanamycin	25 mg/ml in H_2O_{dest}	25 µg/ml
Chloramphenicol	30 mg/ml in EtOH	30 µg/ml
Tetracyclin	10 µg/ml	10 ng/ml
Carbenicillin	50 mg/ml	50 µg/ml
L-Arabinose	500 mg/ml in H_2O_{dest}	0.5-4 mg/ml
IPTG	1 M in H_2O_{dest}	1 mM
X-Gal	40 mg/ml in N,N-DMF	40 µg/ml

Table 3: Inhibitors and medium additives

The *E. coli* strains could be stored on agarplates for a period of up to two months; *Clostridia* cultures could be stored in unopened serum bottles due to their sporulation for six months at 4 °C. For long term storiage “glycerol cultures” were prepared. Cultures were incubated in suitable medium up to the late-logarithmic phase and in each case 2 ml of culture suspension in the relationship 1:1 were mixed with 2 ml sterile glycerol, shock-frozen in liquid nitrogen and stored at -80 °C. Glycerol was added to *C. thermocellum* cultures under anaerobic conditions.

3.3 Molecular genetic methods

3.3.1 Isolation of genomic DNA

For the isolating genomic DNA from *C. thermocellum* 100 ml freshly grown culture in logarithmic grow phase were harvested by centrifugation (5,000 rpm, 20 min, 4 °C; Hettich Rotina 35R, Tuttlingen, Germany). The resulting pellets were resuspended in 5 ml lysis buffer, collected in a 100 ml flask and incubated at room temperature (RT) for 30 min.

Lysis buffer: 20mM Tris pH 7.5
25mM EDTA
75mM NaCl
1mg/ml Lysozyme

After addition of 1/10 (v/v) 10 % SDS and 1 mg/ml Proteinase K (Applichem, Darmstadt, Germany) the flask was slightly shaken for 2 hours. Lysozyme degrades bacterial cell wall hydrolyzing the of 1,4-β-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in peptidoglycan. Due to the treatment with 1 % SDS membrane associated proteins were denaturated. The combined effects result in cell lysis. The mixture was filled up with 1/3 (v/v) 5 M NaCl and 1 volume of chloroform was added to precipitate the proteins. The mixture was slightly agitated for 30 min on a shaker (TR-125 Infors AG, Bottmingen, Switzerland). After centrifugation (4,000 rpm, 12 min) 3 phases (hydrous phase, interphase, and chloroform phase) appear. The DNA containing hydrous phase was collected and mixed with 1 volume isopropanol to precipitate the DNA. After additional centrifugation (5,000 rpm, 10 min) the resulting pellet was washed twice with 70 % ice cold ethanol. The extracted DNA was dissolved in 200-400 µl H_2O_{dest}. The amount of isolated DNA could be determined photometrically (spectrophotometer ND-1000 PEQLab, Erlangen, Germany).

3.3.2 Polymerase chain reaction (PCR)

The PCR was developed in 1983 by Kary B. Mullis (Mullis et al., 1986) and has revolutionized molecular biology founding its way into many disciplines such as basic and applied research medicine, diagnostics and others. The polymerase chain reaction is in general a method for amplification of segments of nucleic acids (up to 10 kb) *in vitro*. The PCR takes advantage of the characteristics of the DNA-polymerases, which extend DNA strands from free 3`-OH-end if a template strand is available. Synthetically made oligodesoxynucleotides of about 20 nucleotides in length (primers) are bound to complementary strands of the target DNA to get 3`-OH-ends, which are elongated. Basically a typical process of PCR is divided into three distinguished temperature steps: denaturation of the template-DNA, primer annealing to a single strand and extension phase.

After denaturation of DNA two primers (forward and reverse) are annealed to the region that contains the section to be amplified. Stable DNA-DNA hydrogen bonds are only formed when the primer sequence very closely matches the template sequence. The polymerase binds to the primer-template hybrid and begins DNA synthesis. By this step the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTPs that are complementary to the template in 5' to 3' direction, condensing the 5'-phosphate group of the dNTPs with the 3'-hydroxyl group at the end of the nascent (extending) DNA strand. At each extension step the amount of DNA target is doubled, leading to exponential (geometric) amplification of the specific DNA fragment.

For optimal conditions of the PCR the right choice of primers is essential. In this work sequence specific primers were used. For their choice some rules have to be noticed:

- at least 17 nucleotides
- average G/C- to A/T-content
- melting temperature T_m ~ 60 °C
- similar T_m for forward and reverse primer
- no hairpin structures (esp. at 3`-end)
- no dimer formation neither with itself nor with second primer
- no G/C-nucleotides at 3`-end to avoid mispriming

- no unusual base sequel like poly-A or long G/C regions

The melting temperature for used primers was approximately calculated by the following formula:

$$T_m\ [°C] = [(G+C) \times 4] + [(A+T) \times 2]$$

All used primers were solved in H_2O_{bidest} at a concentration of 100 pmol/µl and stored at -20 °C.

Name	Sequence	Restriction site
p1 for	tatca**gcatgc**tcttagttgtggctat	Pae I
p1 rev	ataaag**aagctt**tgccaatttctactaccac	Hind III
p6 for	aacaga**gcatgc**aacacctacaacacctg	Pae I
p6 rev	aagtgt**aagctt**gttcggagttatcgtcgg	Hind III
p2-CBM for	aaatgcagc**ggatcc**gattactttgcttgaagtagg	Bam HI
p2-CBM rev	ttcg**gagctc**ctgacggcggtattgtt	Sac I
p3-4 for	aaa**ggatcc**ggttggcagtgtagtacc	Bam HI
p3-4 rev	ttta**aagctt**actgcatccagat	Hind III
p2-CBM-3 for	aaatgcagc**ggatcc**gattactttgcttgaagtagg	Bam HI
p2-CBM-3 rev	ttact**gagctc**gaatcatctgtcg	Sac I
pCBM-3-4 for	aaaaag**ggatcc**acaccaacaaatacagc	Bam HI
pCBM-3-4 rev	ttta**aagctt**actgcatccagat	Hind III
p1-2-CBM-3 for	tatca**gcatgc**tcttagttgtggctat	Pae I
p1-2-CBM-3 rev	ttact**gagctc**gaatcatctgtcg	Sac I
p1-2-CBM-3-4 for	tatca**gcatgc**tcttagttgtggctat	Pae I
P1-2-CBM-3-4 rev	ttta**aagctt**actgcatccagat	Hind III
pBglB-TN for	ctcactct**gcatgc**aaaaagtaaa	Pae I
pBglB-TN rev	cacc**gtcgac**aaacgttcctttaag	Sal I
pCbh9A for	tggtacc**gagctc**cgtgtttgcc	Sac I
pCbh9A rev	ttaaaaagg**cccggg**aaaaaccgg	Sma I
pCipA for	caccaaaagtcatcagt**gagctc**ttagtt	Sac I
pCipA rev	cag**gtcgac**gtaatctcttgatgt	Sal I
pCel9J for	atatct**gcatgc**gccgaaacag	Pae I

pCel9J rev	tttg**cccggg**cttataacttgc	Sma I
pCel9K for	tcaactgttt**gagctc**tggaagac	Sac I
pCel9K rev	atgtcaaccagtaat**gtcgac**tatttttc	Sal I
pCel9R for	ca**ggatcc**tgtttttgcagcagactataac	Bam HI
pCel9R rev	tagctt**gagctc**tttgttttaaagaatacg	Sac I
pCel48S for	aact**gcatgc**gcaggtccttacaaaggc	Pae I
pCel48S rev	aaaagac**ctgcag**aagccgtcc	Pst I
pXynA for	tacc**ggatcc**ctgactttgtttc	Bam HI
pXynA rev	agagaat**cccggg**gcaataaagc	Sma I
pXynC for	tcg**ggatcc**gcggaagtttttt	Bam HI
pXynC rev	aa**cccggg**cttcagccattg	Sma I
pXynZ for	ttggct**ggatcc**ttatgacatcg	Bam HI
pXynZ rev	ggcatcattatctgct**aagctt**ttcg	Hind III
pXghA for	tta**gagctc**ctgtttttgcggc	Sac I
pXghA rev	aga**aagctt**tgcaggttaacacg	Hind III

Tab. 4: used primers, restriction site is in bold print

PCR was carried out with either HotStarTaq Master Mix (Qiagen, Hilden, Germany) for short amplicons up to 1 kb or with *Pfu*-DNA-Polymerase (Fermentas St. Leon-Rot, Germany) for longer fragments. Reactions were performed in a thermocycler PTC-100 (MJ Research Inc., San Fransisco, USA).

Reaction composition for HotStarTaq Master Mix:

Component	Volume/reaction	final concentration
HotStarTaq Master Mix	25 µl	2.5 units/reaction 1 x PCR Buffer 200 µM of each dNTP
Diluted primer forward	2 µl	0.1 – 0.5 µM
Diluted primer reverse	2 µl	0.1 – 0.5 µM
Template DNA	1 µl	< 1 µg/reaction
RNase free water	20 µl	
Final volume	50 µl	

Reaction composition for *Pfu*-DNA Polymerase:

Component	Volume/reaction	final concentration
10 x PCR Buffer	5 µl	1 x PCR Buffer
dNTP mix	5 µl	200 µM of each dNTP
Diluted primer forward	2 µl	0.1 – 0.5 µM
Diluted primer reverse	2 µl	0.1 – 0.5 µM
Template DNA	1 µl	< 1 µg/reaction
Pfu-DNA Polymerase	0.5 µl	1.25 units/reaction
RNase free water	34.5 µl	
Final volume	50 µl	

Amplification program for HotStarTaq Master Mix:

Step	Temperature [°C]	Time [min]	number of cycles
Initiation	95	15	1
Denaturation	94	1	30
Annealing	T_m – 5 °C	1	
Extension	72	1/kb	
Final extension	72	10	1

Amplification program for *Pfu*-DNA Polymerase:

Step	Temperature [°C]	Time [min]	number of cycles
Initiation	95	2	1
Denaturation	95	2	30
Annealing	T_m – 5 °C	1	
Extension	72	2/kb	
Final extension	72	5	1

PCR products were analyzed by agarose gel electrophoresis (3.3.6) and were purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) in accordance with the manufacturer's protocol.

3.3.3 Enzymatic modification of DNA

3.3.3.1 DNA restriction

Endonucleases are enzymes that cleave the phosphodiester bond within a polynucleotide chain, in contrast to exonucleases, which cleave phosphodiester bonds at the end of a polynucleotide chain. Some endonucleases cleavages (*SmaI*) produce blunt ends. Such enzymes are thought to have evolved to provide a defence mechanism against invading viruses. Bacteria prevent their own DNA from being cut by modifying their nucleotides via DNA methylation (Arber et al., 1969).

For digestion of DNA, restriction type II nucleases from prokaryotes were used. Typical type II restriction enzymes differ from type I restriction enzymes in several ways. They are composed of only one subunit; their recognition sites are usually undivided and contain palindromic structures of 4 - 8 nucleotides; they recognize and cleave DNA at the same site; and they do not use ATP for their activity – they usually require only Mg^{2+} as a cofactor (Pingoud et al., 2001). This feature makes this type of enzymes an ideal tool for genetic engineering. The restriction enzyme recognize a specific sequence of nucleotides and produce a double-stranded cut in the DNA by cutting the phosphodiester bond between two bases. Recognition sequences in DNA differ for each restriction enzyme, producing differences in the length, sequence and strand orientation (5' end or the 3' end) of a sticky-end overhang of an enzyme restriction. For cloning of PCR product into vector DNA or controlling insertion of DNA into plasmid the following reaction mixtures were carried out:

Component	Volume/reaction
DNA/vector solution	x µl
10 x reaction buffer	4 µl
Restriction enzyme	3 - 5 units (1 µl)/reaction
Sterile H_2O_{dest}	35 - x µl
Final volume	40 µl

The mixes were incubated at 37 °C (*SmaI* at 30 °C) for 1 – 2 h in a thermo block (AccuBlock Labnet, Woodbridge, USA). For further cloning procedures restriction reactions were purified by QIAquick PCR Purification Kit to remove enzymes, nucleotides and buffer solutions.
All restriction enzymes and reaction buffers were purchased from Fermentas, St. Leon-Rot, Germany.

3.3.3.2 Dephosphorylation of digested vectors

Calf Intestine Alkaline Phosphatase (CIAP, Fermentas St. Leon-Rot, Germany) catalyzes the release of 5`-phosphate groups from DNA, RNA, deoxyribonucleoside and ribonucleoside mon-, di- and triphosphates. If smooth or sticky ends with overlapping sequence are generated during the linearization of the vector, the two ends of the vector can religate and thus reduce the yield of linearized plasmides. By treatment with CIAP religation of the digested vectors is prevented by dephosphorylation of 5`-end of the vector.
After DNA restriction and purification all vectors were treated with CIAP by incubation at 37 °C for 30 min in a thermo block. CIAP can be inactivated by heating at 85 °C for 15 min. For further cloning procedures all reaction mixtures were purified by QIAquick PCR Purification Kit to remove enzymes and buffer solutions.

Component	Volume/reaction
Digested vector solution	x µl
10 x reaction buffer	4 µl
CIAP	1 unit (1 µl)/reaction
Sterile H_2O_{dest}	35 - x µl
Final volume	40 µl

3.3.3.3 Ligation of DNA

The T4 DNA Ligase (Fermentas, St. Leon-Rot, Germany) catalyzes the formation of a phosphodiester bond between juxtaposed 5'-phosphate and 3'-hydroxyl termini in duplex DNA or RNA with blunt or cohesive-end termini. The enzyme repairs single-strand nicks in duplex DNA, RNA or DNA/RNA hybrids but has no activity on single-stranded nucleic acids. The T4 DNA Ligase requires ATP as cofactor. Purified restricted DNA insert and purified restricted plasmid vector, treated with the same restriction enzymes, were ligated overnight at 16 °C in a 20 µL ligation reaction. For best results, for each ligation a molar excess of insert DNA in a ratio of insert to vector 2:1 was used.

Component	Volume/reaction
Digested vector solution	x µl
Digested insert DNA	y µl
10 x reaction buffer	2 µl
T4-DNA-Ligase	5 unit (1 µl)/reaction
Sterile H_2O_{dest}	17 – (x + y) µl
Final volume	20 µl

3.3.3.4 Champion pET TOPO cloning (Invitrogen, Karlsruhe, Germany)

In this system, PCR products are directionally cloned by adding four bases to the forward primer (CACC). The overhang in the cloning vector (GTGG) invades the 5′ end of the PCR product, anneals to the added bases, and stabilizes the PCR product in the correct orientation. Inserts can be cloned in the correct orientation with efficiencies equal to or greater than 90 %. Topoisomerase I from *Vaccinia* virus binds to duplex DNA at specific sites and cleaves the phosphodiester backbone downstream of 5′-CCCTT in one strand (Shuman, 1991). The energy from the broken phosphodiester backbone is conserved by formation of a covalent bond between the 3′-phosphate of the cleaved strand and a tyrosyl residue of topoisomerase I. The phospho-tyrosyl bond between the DNA and enzyme can subsequently be attacked by the 5′ hydroxyl of the original cleaved strand, reversing the

reaction and releasing topoisomerase (Shuman, 1994). This system was used for the expression of the CipA protein.

3.3.4 Transformation of chemical competent *E. coli*

3.3.4.1 Preparation of chemical competent cells

E. coli cells competent for chemical transformation according to the PEG-method were generated from the *E. coli* strains DH5α and M15. Cryo cultures were streaked out on LB plates without antibiotics and incubated overnight at 37 °C. A single colony was suspended in 5 mL LB medium and grown overnight at 37 °C under shaking at 225 rpm. This overnight culture was diluted 1:40 in fresh LB medium and grown to an optical density (OD_{600}) of 0.4 to 0.6 measured with a spectrophotometer (Ultraspec plus Biochrom, Cambridge, England) and centrifuged (5,000 rpm, 10 min, 4 °C). The following steps were performed at 4 °C. Cell pellets were resuspended in 4 ml autoclaved PEG-buffer and subsequently portioned in 100 µL aliquots. The tubes were shock frozen with liquid nitrogen and stored at -80 °C prior use for transformation.

PEG-buffer:
70 ml LB-medium
15 ml Glycerin
10 g PEG 6000
5 ml Dimethyl sulfoxide (DMSO)
50mM $MgCl_2$

3.3.4.2 Chemical transformation

Transformation is defined as the transfer of genetic information into a recipient bacterium using naked DNA without any requirement for contact with a donor bacterium. The ability intake of exogenous DNA (transformation) is generally referred to as competence. Natural competence occurs in a defined subset of bacterial species that have the capacity to take up linear and sometimes circular DNA, usually dependent on a specific uptake system. As natural competence is restricted to a subset of bacteria, methods for the chemical induction of a competent state in otherwise non-transformable bacteria are an important tool in bacterial genetics. For these species, competence refers to the ability to take up and propagate plasmid

DNA, usually with no sequence specificity for uptake. Although not fully understood, chemical methods for the transformation of *Escherichia coli* probably work by transiently opening gated membrane channels. The most common method requires treatment of cells with polyvalent cations and incubation at low temperature. Transient periods of heat and ionic shock probably result in a rapid influx of extracellular medium into the bacterium, after which a recovery period on rich, non-selective medium is usually necessary to ensure full viability of the transformants.

For performing chemical transformation frozen *E. coli* cells were thawed on ice. After the addition of 5 – 10 ng plasmid DNA, the cells were incubated for 30 min on ice. The *E. coli* cells were treated with a heat shock at 42 °C for 30 sec and subsequently cooled down on ice for 2 min. The heat shock initiated the incorporation of DNA-salt complexes into the bacteria. 250 µl of autoclaved SOC-medium was added and cells were shaken with 225 rpm at 37 °C for 1 hour in a thermo block (Thermomixer 5436 Eppendorf, Hamburg, Germany). After incubation, cells were plated onto LB plates supplemented with the appropriate antibiotics and the plates were incubated overnight at 37 °C.

SOC-medium:
2 g Tryptone
0.5 g Yeast-extract
2.5 mM KCl
98 ml H_2O_{dest}
2 ml 1M Glucose soltution sterile filtered

3.3.5 Plasmid DNA preparation

Single transformed *E. coli* colonies were suspended in LB medium supplemented with the appropriate antibiotic. When applying the QIAprep Spin Plasmid Miniprep Kit (Qiagen, Hilden, Germany), 2-5 ml of an overnight culture was sedimented by centrifugation (13,000 rpm, 10 min) and the pellets were used for plasmid DNA preparation in accordance with manufacturer's protocol. This kit is based on the alkaline lysis of *E. coli* and on subsequent absorption of plasmid DNA to silica columns in the presence of high salt concentrations. Plasmid DNA was eluted with 50 – 80 µl H_2O_{dest}. The concentration and purity of plasmid DNA preparations was determined

photometrically with a spectrophotometer (ND-1000 PEQLab, Erlangen, Germany).
For large scale rapid isolation of plasmid DNA the following two alternative methods were performed.

3.3.5.1 Alkaline extraction

This protocol is based on the alkaline lysis of *E. coli* cells. Through a series of steps involving neutralisation, agitation, precipitation, centrifugation, and the removal of supernatant, cellular debris is removed and the plasmid is isolated and purified. Bacteria containing the extrachromosomal plasmid DNA of interest are first cultivated overnight with appropriate antibiotic. Cells in 1.5 ml of culture medium were collected by centrifugation (13,000 rpm, 10 min) and resuspended in 150 µl buffer P1. Cells were lysed with 150 µl of fresh strong alkaline buffer consisting of the detergent sodium dodecyl sulfate (SDS) and the strong base sodium hydroxide. The detergent breaks the phospholipid bilayer of the membrane and the alkali denatures proteins involved in maintaining the structure of the cell membrane. Alkaline conditions were neutralized with 150 µl neutralisation buffer and centrifuged (13,000 rpm, 10 min). Supernatant was collected and DNA precipitated with 500 µl isopropanol. After additional centrifugation (13,000 rpm, 10 min) the precipitate was washed two times with ice cold 70 % ethanol. The washed DNA pellet was dried at room temperature and solved in 40 µl H_2O_{dest}.

Buffer P1:	50 mM Tris 10 mM EDTA 100 µg/ml RNase pH 8
Lysis buffer:	200 mM NaOH 1 % (w/v) SDS
Neutralisation buffer:	60 ml 5 M potassium acetate 28.5 ml glacial acetic acid (99 %) 11.5 ml H_2O_{dest} pH 4.8

3.3.5.2 Plasmid DNA isolation by boiling

The boiling lysis procedure is quick to perform and therefore especially suitable for screening a large number of small-volume *E. coli* cultures. The quality of the isolated plasmid DNA is lower than that of alkaline lysis miniprep (see 3.3.5.1), but it is sufficient for restriction analyses. The bacteria are lysed by treatment with lysozyme, Triton, and heat. 1.5 ml of overnight culture medium containing cells with the target plasmid was centrifuged (13,000 rpm, 10 min) and the resulting pellet was resuspended in 300 µl STET-buffer. After the addition of 10 µl lysozyme solution (10 mg/ml) samples were boiled for 90 sec. The chromosomal DNA, that remains attached to the bacterial membrane and cell debris are removed by centrifugation (13,000 rpm, 15 min). The resulting pellet was removed with a sterile tooth pick. The plasmid remains in the supernatant and is precipitated by the addition of 200 µl isopropanol. Precipitate was pellet by centrifugation (13,000 rpm, 10 min) and washed twice with ice cold 70 % ethanol. After removing ethanol the precipitate was dried at room temperature and solved in 40 µl H_2O_{dest}.

STET-buffer:	8 % (w/v) Saccharose
	5 % (v/v) Triton X 100
	50 mM EDTA
	50 mM Tris-HCl
	5 µg/ml RNase
	pH 8

3.3.6 Agarose gel electrophoresis

To visualize successful PCR reactions or restriction digests, agarose gel electrophoresis was performed. Agarose gel electrophoresis can be used for the separation of DNA fragments ranging from 50 base pair to several megabases using specialized apparatus. The technique of electrophoresis is based on the fact that DNA is negatively charged at neutral pH due to its phosphate backbone. If an electrical potential is applied DNA fragments migrate toward the positive anode and separate by size. The rate at which the DNA move toward the positive pole is slowed by forcing the DNA through an agarose gel. The gel forms a porous lattice and larger molecules are slowed down more than smaller molecules, since the smaller molecules can

fit through the mesh more easily. As a result, a mixture of large and small fragments of DNA will be separated by size. The distance between DNA bands of a given length is determined by the percentage of agarose in the gel. Most agarose gels are prepared with between 0.7 % (large 5–10 kb DNA fragments) and 2 % (small 0.2–1 kb DNA fragments) agarose dissolved in electrophoresis buffer. 1 % gels are common for many applications. The most common dye used to visualize DNA or RNA bands for agarose gel electrophoresis is ethidium bromide. When exposed to ultraviolet light, it will fluoresce with an orange colour, intensifying almost 20-fold after intercalating with DNA or RNA.

1 % (w/v) Agarose in TAE buffer was dissolved by boiling in a microwave. The agarose solution was poured into gel trays and the combs were inserted. After solidification of the gels, the combs were removed and the trays were put into the electrophoresis chamber. The gels were overlaid with TAE buffer. DNA samples were treated with gel loading buffer and loaded onto the gel. Electrophoresis was performed with 80-100 V using a power supply until the bromophenol blue dye front reached the end of the gel. The gel was stained in a 0.01 % ethidium bromide water bath for 10 minutes and DNA fragments were visualized by an UV transilluminator (Biostep, Jahnsdorf, Germany). To determine the size of the DNA fragments a 10 kb DNA standard (GeneRuler DNA ladder mix Fermentas, St. Leon-Rot, Germany) was used. For purification, DNA fragments were excised with a scalpel and eluted from the gel plugs by the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) in accordance with the manufacturer's protocol.

50 x TAE-buffer:	2 M Tris 100 mM EDTA 1 M acetic acid pH 8.5
10 x loading dye:	50 % (v/v) Glycerin 0.5 M EDTA 0.25 % (w/v) bromophenol blue 0.25 % (w/v) Xylenecyanol

3.4 Protein biochemistry

3.4.1 Screening for positive clones

Rapid plasmid preparation, restriction analyses and gel electrophoresis were performed with DNA from clones with inserts that encode for non-hydrolytic active enzyme. Positive clones overexpressing recombinant hydrolytic enzymes were detected by Congo Red staining. Congo Red has a strong though apparently non-covalent affinity to polymeric glucans, including cellulose. The soluble barley β-glucan, a β-1,3-1,4-glucan, can be visualized by staining with Congo Red however, short chained degraded products like cellobiose or cellotetrose generated by enzymatic hydolysis of barley β-glucan do not interact with Congo Red. Single colonies were picked and streaked out on agar plates and additionally on parallel plates. After incubation over night the plates heated on 60 °C for 30 min were overlayed with molten soft agar (0.8 %) containing barley β-glucan. The reaction mixtures were incubated at 60 °C for several hours. By adding 0.5 % Congo Red solution to the plates, incubating for 10 min and additionally destaining with 1 M NaCl for 15 min, the cellulose in the overlay was stained. However, clear halos appeared around colonies producing recombinant hydrolytic enzymes for cellulose degradation.

Agarose overlay:	0.8% (w/v) agarose 0.06% (w/v) Barley β-glucan 50 mM Tris-HCl

3.4.2 Expression of recombinant proteins

High-level expression of 6xHis-tagged proteins in *E. coli* using pQE vectors is based on the T5 promoter transcription-translation system. These plasmids contain an optimized promoter–operator element consisting of the phage T5 promoter that is recognized by the *E. coli* RNA polymerase, and two *lac* operator sequences. The extremely high transcription rate initiated at the T5 promoter can only be efficiently regulated and repressed by the presence of high levels of the *lac* repressor protein. The *E. coli* M15 host strain used in the pQE system uses a *lac* repressor gene in *trans* to the gene to be expressed. The host strain contains the low-copy plasmid pREP4, which

confers kanamycin resistance and constitutively expresses the *lac* repressor protein encoded by the *lac* I gene (Farabaugh, 1978). The pREP4 plasmid is derived from pACYC and contains the p15A replicon. Multiple copies of pREP4 are present in the host cells that ensure the production of high levels of the *lac* repressor protein which binds to the operator sequences and tightly regulates recombinant protein expression. The pREP4 plasmid is compatible with plasmids carrying the ColE1 origin of replication, and is maintained in *E. coli* in the presence of kanamycin. Expression of recombinant proteins encoded by pQE vectors is rapidly induced by the addition of isopropyl-β-D-thiogalactoside (IPTG), which binds to the *lac* repressor protein and inactivates it. Once the lac repressor is inactivated, the host cell's RNA polymerase can transcribe the sequences downstream of the promoter.

Additionally an alternative expression system was used in this work. The pET vector is based on the T7 promotor transcription-translation system. The pET101/D-TOPO system (Invitrogen, Karlsruhe, Germany) allows expression of recombinant protein with a native N-terminus and a C-terminal fusion tag. The Champion pET Expression System uses elements from bacteriophage T7 to control expression of heterologous genes in *E. coli*. In the pET TOPO vectors, expression of the gene of interest is controlled by a strong bacteriophage T7 promoter that has been modified to contain a *lac* operator sequence. In bacteriophage T7, the T7 promoter drives expression of gene 10 (φ10). T7 RNA polymerase specifically recognizes this promoter. To express the gene of interest, it is necessary to deliver T7 RNA polymerase to the cells by inducing expression of the polymerase. T7 lysozyme binds to the T7 RNA polymerase and inhibits transcription. This activity results in reduced basal levels of T7 RNA polymerase, leading to reduced basal expression of T7-driven heterologous genes. In addition to the gratuitous inducer, isopropyl β-D-thiogalactoside (IPTG) allows expression of T7 RNA polymerase from the *lacUV5* promoter of the pLysSRARE plasmid. The Rosetta-gami B (DE3)pLysS cells (Novagen, Darmstadt, Germany) contain the pLysSRARE plasmid, which produces T7 lysozyme and contains additional sequences for production of rare t-RNAs.

A 50 ml pre-culture of *E. coli* with the plasmid of interest was expanded in 500 ml LB-medium containing the appropriate antibiotics. Cells were grown at 37 °C until OD_{600} of 0.4 – 0.6 was reached. Induction of expression took place after the addition of IPTG to a final concentration of 0.1 to 1 mM. After 4 – 6 h incubation at 37 °C (or 30 °C) the cells were harvested for further treatments (see 3.4.6).

3.4.3 Coexpression of chaperones

Expression of foreign proteins in *E. coli* often results in problems, such as the formation of inclusion bodies or protease degradation of the recombinant protein which are frequently encountered in a protein functional research. These issues often are a result of improper folding of the expressed proteins. Molecular chaperones are involved in the protein folding process. The chaperone plasmid pG-KJE8 (Takara Bio Saint-Germain-en-Laye, France) is designed to enable efficient expression of multiple molecular chaperones (dnaK, dnaJ, grpE, groES, groEL) known to work in cooperation in the folding process. The coexpression of a target protein with these chaperones often increases recovery of expressed proteins in the soluble fraction (Nishihara, 1998, 2000).

Before transformation of plasmid encoding the target protein plasmid pG-KJE8 was transformed into chemical competent *E. coli* Top10F` cells. The cells were prepared to be chemically competent (see 3.3.4.1) again and transformed with the plasmid containing the gene of interest. To perform coexpression, transcription of chaperones was induced by adding 2 mg/ml L-arabinose and 5 ng/ml tetracycline after inoculation. The chaperone genes are situated down stream of either the *araB* or *Pzt-1* (tet) promotor. When OD_{600} reached 0.4 – 0.6 IPTG to a final concentration of 0.1 to 1 mM was added and the procedure went on as previously described (see 3.3.7.2).

3.4.4 Sonification of cells

This procedure is used to break cells that contain a protein to be purified. Intensive treatments with ultrasound break the cell walls and shear the DNA into sizes that will not affect the viscosity of the samples. Harvested cells after induction with IPTG were resuspended in 20 ml MOPS-buffer (50mM MOPS, 0.1 M NaCl, 5mM $CaCl_2$, 20 mM imidazol, pH 7.3) and incubated with 1 mg/ml lysozyme on ice for 30 minutes. Cells were filled in a glass vial and placed in a container with ice water and salt (NaCl). The reaction mixture was treated with an ultrasonicator UP200S (Dr. Hielscher GmbH, Teltow, Germany) with an amplitude of 75% and short pulses for 5 minutes. After 5 minutes of cooling, the sonification procedure was repeated. Cell suspension was centrifuged with a high speed centrifuge (18,000 rpm, 20 min, Sorvall RG5B Plus, Thermo Scientific, Langenselbold, Germany) to settle cell debris. The resulting supernatant contains the proteins to be purified and was treated with a protease inhibitor cocktail complete mini (Roche, Mannheim, Germany).

3.4.5 Purification of inclusion bodies

When genes from one organism are expressed in another the resulting protein sometimes forms inclusion bodies. Inclusion bodies are nuclear or cytoplasmic aggregates of proteins. Many recombinant polypeptides are unable to fold properly within the cell and associate to form large protein aggregates. Indirect evidence suggests that the conformation of different proteins in inclusion bodies can vary from a native like state to completely misfolded molecules that are dissociated only under stringent denaturing conditions. Inclusion bodies can be detected by light microscopy. They form dark spots primarily at the poles within the cell.

To purify inclusion bodies and refold the aggregated, misfolded proteins an *E. coli* culture was disrupted by sonification after induction with IPTG. Cells were harvested by centrifugation (15,000 rpm, 20 min) and the resulting pellet was resuspended in 100 mM Tris-HCl, 5 M Urea, pH 8.8. Inclusion bodies are now dissolved under denaturing conditions. Supernatant was collected after additional centrifugation (15,000 rpm, 20 min) and dialysed against 20 mM Tris-HCl, 1.5 mM Cellobiose, pH 7.0 at 4 °C using a Slide-A-

Lyzer cassette (cutoff of 10,000 MW, Pierce, Bonn, Germany) with at least 4 buffer exchanges within 30 hours. The content of the cassette was removed and collected for further purification steps (see 3.4.6).

3.4.6 Purification of recombinant His-tagged proteins via affinity chromatography

A prerequisite for protein purification is a method for detection due to their biochemical characteristics in a protein mixture. The FPLC (Fast Protein Liquid Chromatography) is a form of liquid chromatography with medium pressure (maximal 5 MPa). A fast separation of the proteins is ensured by the use of chemically-physically stable and biocompatible column materials with high resolution and high reproducibility.

If histidine side chains of a protein are positioned spatially neighbouring to each other, then their imidazol rings will posses a high affinity to bivalent ions such as Ni^{2+}. With recombinant expression a protein can be fused with the six-fold repeated amino acid histidine and an artificial connection side for the affinity chromatography at Ni-NTA-Agarose is produced. Consequently, a protein containing a histidine tag will be selectively bound to metal-ion-charged media such like sepharose (Ni-NTA-Agarose) while other cellular proteins will not bind or bind weakly. This chromatographic technique is often termed immobilized metal ion affinity chromatography (IMAC). In general, the histidine-tagged protein is the strongest binder among all the proteins in a crude sample extract from, for example, a bacterial lysate. Imidazole competes with proteins for binding to Ni^{2+} sepharose column and an excess of imidazole is passed through the column, which displaces the His-tag from nickel co-ordination, freeing the His-tagged proteins.

For purification of recombinant His-tagged proteins the system from Pharmacia (Freiburg, Germany) containing the following components was used: Gradient-programmer GP 250, double channel pump P500, manual injection device V7 with loops to 50 ml, UV-monitor Uvicord S II (λ = 280 nm), chart recorder REC 102

The column HisTrap FF 5ml (GE Healthcare Munich, Germany) was equilibrated with 5 volumes of washing buffer. The cell-free extract

containing tagged proteins was applied into the column with a flow rate of 2 ml/min. By additional washing with at least 5 volumes of washing buffer unbound proteins were removed from column until the absorbance reached the steady baseline. Tagged proteins were eluated with elution buffer and different elution fractions were collected. Regeneration and cleaning procedure of column took place in accordance with the manufacturer's protocol.

Washing-buffer:	50 mM MOPS 0.1 M NaCl 20 mM imidazole 5 mM $CaCl_2$ pH 7.2
Elution-buffer:	50 mM MOPS 0.1 M NaCl 0.5 M imidazole 5 mM $CaCl_2$ pH 7.4
Stripping-buffer:	50 mM MOPS 0.1 M NaCl 50 mM EDTA pH 7.2

3.4.7 Purification of native cellulosome

Cellulosomes were isolated from cell-free culture broth using the affinity digestion method adapted by Zhang et al. (2003). 1 l from a *Clostridium thermocellum* culture was spun down and the cell free supernatant was incubated with 100 mg/l phosphoric acid swollen cellulose (PASC) overnight at 4 °C for cellulosome binding to cellulose. Amorphous cellulose with bound complexes was collected by centrifugation (15000 rpm, 15 min, 4°C) and resuspended in 20 ml dialysis buffer (50 mM Tris, 5mM $CaCl_2$, 5mM DTT, pH 7.0). The suspension was dialyzed in a Slide-A-Lyzer cassette (cutoff 10,000 MW) at 60 °C against 2 l of H_2O_{dest} to remove produced small molecules through the initiated degradation of amorphous cellulose by the hydrolytic complexes. Deionized water was changed every hour to avoid inhibition of cellulosome activity by the degradation product cellobiose. The

suspension cleared within 5-6 hours and purified cellulosome fraction was obtained after further centrifugation (13,000 rpm, 15 min) of the solution.

3.4.8 Purification of SM901mutant extracellular proteins

A well grown culture of *Clostridium thermocellum* mutant SM901 (Zverlov et al., 2008) was spun down by centrifugation (13,000 rpm, 20 min). Saturated $(NH_4)_2SO_4$ solution was added to the supernatant to a final concentration of 60 % to precipitate extracellular enzymes and incubated overnight at 4°C. Proteins were collected by centrifugation (15,000 rpm, 20 min, 4 °C) and resuspended in 50 mM MOPS, 0.1 M NaCl, 5 mM $CaCl_2$, 0.02% NaN_3, pH 7.0.

3.4.9 SDS polyacrylamide gel electrophoresis (PAGE)

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis, is a technique to separate proteins according to their electrophoretic mobility. The SDS gel electrophoresis of samples having identical charge to mass ratios results in fractionation by size. The samples are boiled with SDS (sodium dodecyl sulfate), an anionic detergent which gives negative charge to each protein in proportion to its mass. SDS linearizes the proteins so that they may be separated strictly by molecular weight. In order to separate proteins from sample volumes of 110 to 50 µl, Laemmli developed the discontinuous SDS gel electrophoresis in 1970. DISC electrophoresis utilizes gels of different pore size, a stacking gel and a separation gel.

The stacking gel has the task of concentrating the proteins at the front to the separation gel. In the stacking gel a pH value of 6.8 is present. Under these conditions glycine molecules of running buffer are present as zwitterions (subsequent ions) while the chloride ions run ahead (guidance ions). This causes a lack of charge carriers. Thus a zone of high electrical field strength exists between the guidance ions and the subsequent ions. The SDS coated proteins arrange themselves between guidance and subsequent ions, and migrate propelled from the increased field strength towards the separation gel front. Thus the proteins are concentrated at the separation gel front. This improves separation of the bands substantially.

Separation by molecular weight takes place in the small porous separation gel. If the SDS coated proteins of the stacking gel cross into the separation gel, they are slowed down by the filter effect of the narrow pores. As soon as the glycine ions of the collecting gel migrate into the basic separation gel (pH 8.8), they receive a negative net charge and overtake the macromolecular anions. The lack of charge carriers is waived and the proteins are isolated after their molecular weight.

Stacking gel and separation gel are prepared by the following schema.

Stacking gel buffer:	0.5 M Tris 0.4 % (w/v) Sodium dodecyl sulfate pH 6.8 with HCl
Separation gel buffer:	1.5 M Tris 0.4 % (w/v) Sodium dodecyl sulfate pH 8.8 with HCl

Stacking gel (3%)

Component	Volume/2 gels
H_2O_{dest}	1.625 ml
30% Acrylamide+ 0.8 % N,N`-Methylenbisacrylamide	0.25 ml
Stacking gel buffer	0.625ml
TEMED	8 µl
Ammonium persulfate (10 %)	15 µl
Final volume	2.52 ml

Separation gel (10 %)

Component	Volume/2 gels
H_2O_{dest}	2.15 ml
30% Acrylamide + 0.8 % N,N`-Methylenbisacrylamide	1.65 ml
Separation gel buffer	1.25 ml
TEMED	15 µl
Ammonium persulfate (10 %)	25 µl
Final volume	5.05 ml

The separation gel solution was cast between two spacer-separated glass plates until filling height reached ~1.5 cm below end of plates. Gel was covered with 0.5 ml H_2O_{dest} to avoid drying-out and incubated at room temperature for 30 minutes. After polymerization the probe well was introduced and separation gel was covered with stacking gel.
Protein probes were mixed with loading dye and incubated in a boiling water bath for 6 minutes. Maximally 20 µl (~10 µg protein/lane) of the probes were applied into the preformed bags of stacking gel. PageRuler prestained protein ladder (Fermentas, St. Leon-Rot, Germany) was used as marker. Electrophoresis was performed in 1x running buffer with the vertical gel electrophoresis system Mini-Protean II (Biorad, Munich, Germany) for 1.5 hours at 20 mA. When bromophenol blue band of loading dye reached the end of the gel, electrophoresis was stopped.

Running buffer (10x):	0.25 M Tris 1.92 M Glycine 1 % (w/v) SDS pH 8.3
Loading dye (4x):	6 ml Glycerine 2 ml β-Mercaptoethanol 1 g SDS 0.5 ml Bromophenol Blue 0.32 g Tris Add H_20_{dest} to 20 ml pH 6.8 with HCl

Staining of gels was performed in the staining solution for 15 minutes at 60 °C. Non-specifically bound dye was removed by incubation in destaining solution for 15 minutes at 60 °C.

Staining solution:	1.5 g Coomassie Brillant Blue 455 ml Methanol 80 ml Glacial acetic acid Add H_2O_{dest} to 1 l
Destaining solution:	250 ml Methanol 350 ml Glacial acetic acid Add H_2O_{dest} to 5 l

3.4.10 Native gel electrophoresis

Native or non-denaturing gel electrophoresis is run in the absence of charged denaturing agents such as SDS. While in SDS-PAGE the electrophoretic mobility of proteins depends primarily on their molecular mass, in native PAGE the mobility depends on both the protein's charge and its hydrodynamic size. The electric charge driving the electrophoresis is governed by the intrinsic charge on the protein at the pH of the running buffer. This charge will depend on the amino acid composition of the protein as well as post-translational modifications such as addition of sialic acids. Since the protein retains its folded conformation, its hydrodynamic size and mobility on the gel will also vary with the nature of this conformation. If native PAGE is carried out near neutral pH to avoid acid or alkaline denaturation, then it can be used to study conformation, self-association or aggregation, and the binding of other proteins or compounds. In this work native gel electrophoresis was performed to study the binding abilities of the protein carriers (miniscaffoldins) to dockerin bearing hydrolases.

Except for heating the probes and the buffer and loading dye composition the procedure of native gel electrophoresis is equal to that of SDS-PAGE (see 3.4.9).

Stacking gel buffer:	0.5 M Tris pH 6.8 with HCl
Separation gel buffer:	1.5 M Tris pH 8.8 with HCl
Running buffer (10x):	0.25 M Tris 1.92 M Glycine pH 8.3
Loading dye (4x):	6 ml Glycerine 0.5 ml Bromophenol Blue 0.32 g Tris Add H_20_{dest} to 20 ml pH 6.8 with HCl

3.4.11 Western blot

The Western blot is an analytical technique used to detect specific proteins in a given sample. The proteins are then transferred to a membrane, where they are detected using antibodies specific for the target protein. Proteins are transferred using an electric current from a SDS-gel (see 3.4.9) into the membrane and bind to the membrane upon hydrophobic interactions, as well as charged interactions between the membrane and the protein.

In this study SDS-protein complexes separated by SDS-PAGE were transferred electrophoretically onto Polyvinylidene difluoride (PVDF) membranes. For immunodetection of individual proteins specific antibodies (anti-His) were used. Proteins in SDS-PAGE slabs were renaturated by incubation in 25 % (v/v) isopropanol, 50 mM Na-citrate for 20 min and washing with 50 mM Na-citrate for 20 min. SDS gels were incubated in transfer buffer for 5 min. PVDF membranes were activated with methanol, rinsed with H_2O_{dest} and incubated in transfer buffer for 5 min. Cut up filter paper (Whatman, Dassel, Germany) was soaked in transfer buffer prior to use. A layer of five filter paper sheets and the PVDF membrane were laid on the anodal side of the blotting unit. The SDS gel was then placed air bubble-free on top of the membrane, followed by a layer of five filter paper sheets. The proteins were electroblotted to the membrane at 1 mA/cm^2 for 2 h using a 2117 Multiphor II Electrophoresis unit (Pharmacia LKB Uppsala, Sweden). The membrane was blocked overnight at 4 °C with 200 g/liter skim milk powder in phosphate-buffered saline buffer and washed with washing buffer for 20 min. To detect cohesin-containing proteins on the blot, a recombinant protein containing a dockerin and a His-tag (from clone rCel9N, *C. thermocellum* Cel9N cellulase) was purified from 400 ml of cell extract of *E. coli* containing *celN* gene bearing pQE-vector. 50 µg of rCel9N per ml per cm^2 of membrane surface was applied for 2 h. After the washing buffer was changed three times, the attached rCel9N protein was detected with peroxidase conjugated anti-His antibodies (Qiagen Hilden, Germany; 1:1,000 in phosphate-buffered saline buffer). The rCel9N/antibody complex was detected with 4-nitroblue tetrazolium chloride and bromochloro-indolylphosphate (Roche Diagnostics GmbH, Mannheim, Germany) in accordance with the manufacturer's recommendations.

Transfer buffer:	50 mM Tris 40 mM Glycine 1 mM SDS 20 % (v/v) Methanol
Washing buffer:	50 mM Tris pH 7.7 0.05 % (v/v) Tween 20 0.5 % (w/v) Bovine serum albumin 2 mM $CaCl_2$
Phosphate buffered saline:	0.1 M NaH_2PO_4 150 mM NaCl pH 7.2

3.4.12 Determination of protein concentration

The concentration of proteins in extracts was determined in accordance with Bradford (1976). This assay is based on the binding of the dye Coomassie Brilliant Blue G-250 to proteins leading to a shift in the absorption maximum of the dye from 465 nm to 595 nm. This shift is presumably caused by the stabilization of the dye in its un-protonated, anionic sulfonated form by complex formation between dye and protein. The dye binds non-specifically to cationic and non-polar, hydrophobic side chains of proteins (mainly arginine and aromatic amino acids). The Bradford method shows a high sensitivity (1 µg protein/mL reaction solution) and is not disturbed by reducing agents such as dithiothreitol (DTT) and β-mercaptoethanol. In contrast, the measurement is compromised by detergents such as sodium dodecyl sulfate (SDS), strong basic reagents and high urea concentrations. The Coomassie Blue P-250 Kit (Pierce Bonn, Germany) was used for the Bradford assay. 0.75 ml of dye solution was mixed with 25 µl of protein sample and incubated for 5 min at RT. Dye reagent mixed with 25 µl H_2O_{dest} was used as blank. Absorption was measured photometrically at 595 nm. Protein concentration was calculated using a standard regression curve (mx + t) with bovine serum albumin (BSA) as protein standard (0-40 µg BSA; see appendix) using the following equation.

$$C_{protein}[mg/ml] = \frac{OD_{595} - t}{m \times V[ml]}$$

3.4.13 Determination of enzyme activity

By splitting of glycosidic bonds new molecule ends are generated in glucan molecule, which affect due to their half acetal character the reducing of oxidative agents. This increase in reduction strength is quantitatively determined in the DNSA test (Miller, 1959), as the nitro-group is stoichiometrically reduced by 3,5-Dinitrosalycylic acid (DNSA) during heating, whereby a color change takes place from yellow to brown.

DNSA-reagent:
10 g 3,5-Dinitrosalycylic acid
2 g Phenol
0.5 g Na_2SO_4
200 g $KNaC_4H_4O_6 \cdot 4\ H_2O$
10 g NaOH
Add H_2O_{dest} to 1000 ml

For determination of hydrolytic activity with soluble and insoluble substrate the following reaction mixtures were carried out.

Component	Volume/reaction
1M MES buffer pH 6.0	75 µl
1 % (w/v) substrate solution	375 µl
Enzyme solution	x µl
0.1 M $CaCl_2$ solution	37.5 µl
H_2O_{dest}	262.5 µl –x µl
Final volume	750 µl

After incubation at different temperatures and for different time periods the samples were centrifuged with 13000 rpm for 10 minutes. 500 µl of the supernatant was mixed with 750 µl of DNSA-reagent to stop the hydrolytic reaction and the samples were boiled for 15 minutes in a water bath. Absorption was measured at 575 nm with a photometer. One enzyme unit is defined as the amount of the enzyme releasing 1 µmol of glucose per min. Specific activity [U/mg] was determined using a standard regression curve with glucose (0 – 500 µg) as standard (see appendix) and calculated with the following equation. Xylanase activity was determined in the same way using

a standard regression curve with xylose (0 – 200 µg) as standard and calculated with molecular mass of xylose (150.13 g/mol) instead of glucose.

$$Specactivity[U/mg] = \frac{Glu[\mu g] \times 1000}{180 \times t[\min] \times V[\mu l] \times c[mg/ml]}$$

An alternative method for determing the release of glucose molecules from the cellulose was applied by incubating the samples in addition to cellulase with β-glucosidase from *T. neapolitana*, to degrade all cellodextrins to glucose. The method is based on the following coupled enzymatic reaction:

$$\text{Glucose} + H_2O \xrightarrow{\text{Glucose Oxidase}} \text{Gluconic acid} + H_2O_2$$

$$H_2O_2 + \text{o-Dianisidine (colorless)} \xrightarrow{\text{Peroxidase}} \text{Oxidized o-Dianisidine (brown)}$$

The intensity of the brown colour measured at 425 nm is proportional to the glucose concentration. PGO solution (Peroxidase – Glucose Oxidase) was prepared by adding one PGO enzymes capsule (Sigma Aldrich, St. Louis, USA) to 100 ml deionised water. The o-dianisidine solution was prepared by dissolving 50 mg of o-dianisidine in 20 ml H_2O_{dest}. The Reaction solution was obtained by mixing 100 ml of PGO solution with 1.6 ml o-dianisidine solution. 100 µl of the glucose containing sample was added to 1 ml of PGO enzyme reaction solution. After incubation at 37 °C for 30 minutes absorption was measured photometrically at 425 nm. The amount of glucose [µg] was determined using a standard regression curve with glucose (0 – 500 µg) as standard (see appendix).

3.4.14 Detection of hydrolytic activity by SDS-gel electrophoresis

To detect hydrolytically active proteins in SDS-gel slabs 0.1 % (w/v) β-glucan was added to the gel solution before polymerisation. The polymer could be degraded by the enzymes resulting in a clear halo after staining the gel with Congo Red. To render the enzymes in the renaturing SDS-gel enzymatically active they must be renaturated by washing out SDS from the gel. Removal of SDS was carried out by incubating the gel slabs in renaturation buffer containing isopropanol with an additional washing step in washing buffer for 20 minutes each. Renaturation and washing steps were repeated two times at room temperature.

Renaturation buffer:	50 mM Sodium citrate 25 % (v/v) Isopropanol pH 6.0
Washing buffer:	50 mM Sodium citrate pH 6.0

After renaturation the gel was incubated in washing buffer at 60 °C for 2 hours to allow enzyme activity. Halos in the gel caused by hydrolysis of β-glucan were detected by shaking the gel slab in 0.5 % (w/v) Congo Red solution for 10 minutes and additional washing with 1M NaCl for 20 minutes. After photographical documentation the stained gel could be stored in 0.1 M Tris buffer at pH 9.

3.4.15 Thin layer chromatography

Thin layer chromatography (TLC) was used for separating mixtures of monomeric and oligomeric sugars. It was performed on a sheet of aluminium coated with the absorbent material silica gel (stationary phase). After the sample has been applied on the plate, the solvent (mobile phase) is drawn up the plate via capillary action. Since different components ascend the TLC plate at different rates, separation is achieved.

First samples were prepared according to determination of hydrolytic activity (see 3.4.14). 5 - 10 µl Samples (~ 10 µg) of reaction probes and standard (0.1 % (w/v) glucose, cellobiose, cellotriose, cellotetrose or xylose, xylobiose, xylotriose, xylotetrose, xylopentose) were applied on a silica gel 60 plate (20 x

20 cm; Merck, Darmstadt, Germany) with a 1.5 cm distance from the bottom. Probes were dried with a warm air flow. The plate was put in a glass chamber filled with running buffer and incubated for 45 to 60 minutes until solvent front reached the top of the plate. After a short period of drying the plate was spread with the detection reagent and incubated at 130 °C in a heater for 10 minutes. Sugars could be detected as dark spots on a yellowish background.

Running buffer:	80 % (v/v) Acetonitrile
Detection reagent:	20 ml stock solution + 2 ml phosphoric acid
Stock solution:	100 ml Acetone 1 ml Aniline 1 g Diphenylamine

3.4.16 Gel filtration (Size exclusion chromatography)

Gel filtration chromatography separates proteins, peptides, and oligonucleotides on the basis of size. Molecules move through a bed of porous beads, diffusing into the beads to greater or lesser degrees. Smaller molecules diffuse further into the pores of the beads and therefore move through the bed more slowly, while larger molecules enter less or not at all and thus move through the bed more quickly. Both molecular weight and three-dimensional shape contribute to the degree of retention. Gel filtration chromatography may be used for analysis of molecular size, for separation of components in a mixture, or for salt removal or buffer exchange from a preparation of macromolecules.

For this work gel filtration was used to separate the proteins of interest from protein mixtures. Superdex 200 10/300 (GE Healthcare, Munich, Germany) with an optimal separation range between 3.000 and $6x10^5$ Mr was used as column material. Initially, the column was equilibrated with two column volume of 50 mM MOPS, 0.5 M NaCl, 30 mM Imidazol, 10 mM $CaCl_2$, pH 7.3. Maximal sample volume was 0.5 ml and flow rate 0.5 ml/min. Fractions of 250 µl were collected with fractionizer Pharmacia LKB FRAC-100 (Amersham Pharmacia, Freiburg, Germany).

3.5 Nanoparticles and coupling chemistry

3.5.1 Nanoparticles

Recently, considerable research has focused on iron oxides due to their potential uses such as pigment, magnetic drug targeting, magnetic resonance imaging for clinical diagnosis, recording material and catalyst. The magnetic nanoparticles exhibit superparamagnetic behavior caused by the infinitely small coercivity arising from the negligible energy barrier in the hysteresis of the magnetization loop of the particles as predicted by Bloch and Neel (Hartmann and Mende, 1986).

The nanoparticles (NP) used in this study were purchased from Estapor Microspheres (Fontenay, France). They have a mean diameter of ~100 nm and the iron oxide core imparts the particles a superparamagnetic character (Fig. 13). The core is covered with a polymeric shell of polystyrol and the surface of the particles is modified either with free carboxyl groups (COOH = 497 µeq/g NP) or with free amino groups (NH_2 = 11 µeq/g NP) for coupling reaction with crosslinking agents.

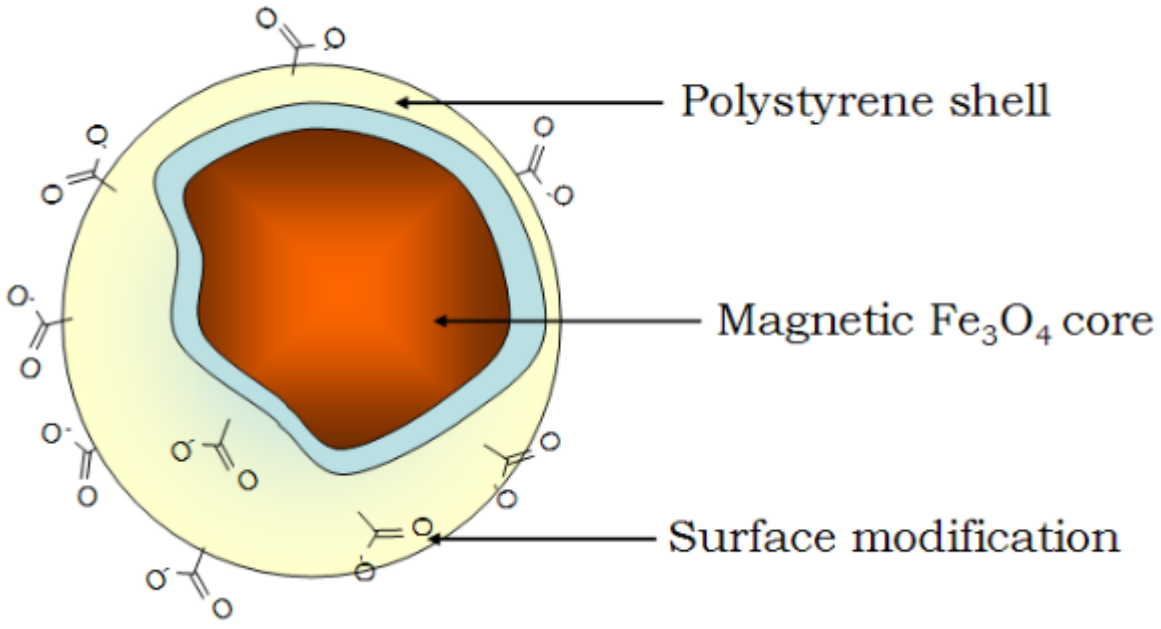


Fig. 13: Structure of used nanoparticles. A superparamagnetic iron oxide core is covered with several layers of polymeric polystyrene. For coupling proteins on nanoparticles the surface is modified with suitable crosslinking molecules such as carboxyl groups.

3.5.2 Coupling chemistry

3.5.2.1 COOH-modified nanoparticles

To immobilize proteins on the surface of nanoparticles the functional groups (free COOH-groups) have to be activated. The water-soluble carbodiimide 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, Pierce, Rockford, Ireland) was used to form active ester groups with carboxylate groups using the water-soluble compound *N*-hydroxysulfosuccinimide (sulfo-NHS, Pierce, Rockford, Ireland). EDC reacts with a carboxylate group to form an active ester (O-acylisourea) leaving group. Sulfo-NHS esters are hydrophilic active groups that react rapidly with amines on target molecule (Staros et al., 1986). However, in the presence of amine nucleophiles that can attack at the carbonyl group of the ester, the sulfo-NHS group rapidly leaves, creating a stable amide linkage with the amine. The advantage of adding sulfo-NHS to EDC is to increase the stability of the active intermediate, which ultimately reacts with the attacking amine. The reaction of EDC with carboxylate groups is subject to rapid hydrolysis in aqueous solution, having a rate constant measured in seconds. Forming a sulfo-NHS ester intermediate from the reaction of hydroxyl group on sulfo-NHS with the EDC active-ester complex extends the half-rate of the activated carboxylate to hours (Hermanson, 1995).

20 mg of carboxyl-modified nanoparticles were washed three times in 2 ml activation buffer (50 mM MES, 0.5 M NaCl, pH 6) in a glass vial by separation with a strong NdFeB disc magnet (1.41 – 1.45 Tesla). The modified surface of particles was activated by adding fresh EDC solution and sulfo-NHS solution to a final concentration of 2 mM and 5 mM, respectively (see Fig. 14). The mixes reacted for 15 minutes at RT. Particles were separated with magnet and washed two times with 2 ml reaction buffer (0.1 M sodium phosphate, 0.5 M NaCl, pH 7.2). 5 mg of O-(2-aminoethyl)-O-(2-carboxyethyl)-polyethylenglycol 3000 hydrochloride (NH_2-PEG-COOH) was dissolved in 100 µl reaction buffer under nitrogen atmosphere and added to the activated particles. The covalent link between the activated particles and the amino groups of the PEG based linkers took place within 3 hours at RT. Buffer was changed to 2 ml activation buffer and the carboxylate group at

the end of the covalently bound linker was activated with EDC and sulfo-NHS as previously described. After two washing steps with 2 ml reaction buffer 10 mg N_α, N_α-Bis(carboxymethyl)-L-lysine Hydrate (NTA) was added. Coupling of NTA to activated carboxylate groups of linker took place within 3 hours at RT. Particles were washed three times with 2 ml distilled water and 1 ml of 1M $NiSO_4$ was added. The free Ni^{2+} ions were complexed by carboxylate groups of NTA and NTA-Ni was formed. After 5 minutes the particles were washed two times with 2ml distilled water and additionally two times with 2 ml 50 mM MOPS, 0.1 M NaCl, 5 mM $CaCl_2$, pH 6. Conjugation of His-tagged protein carriers with nanoparticles was achieved by incubating the particles with 1 to 1.5 mg protein overnight in 2 ml 50 mM MOPS, 0.1 M NaCl, 5 mM $CaCl_2$, pH 6.

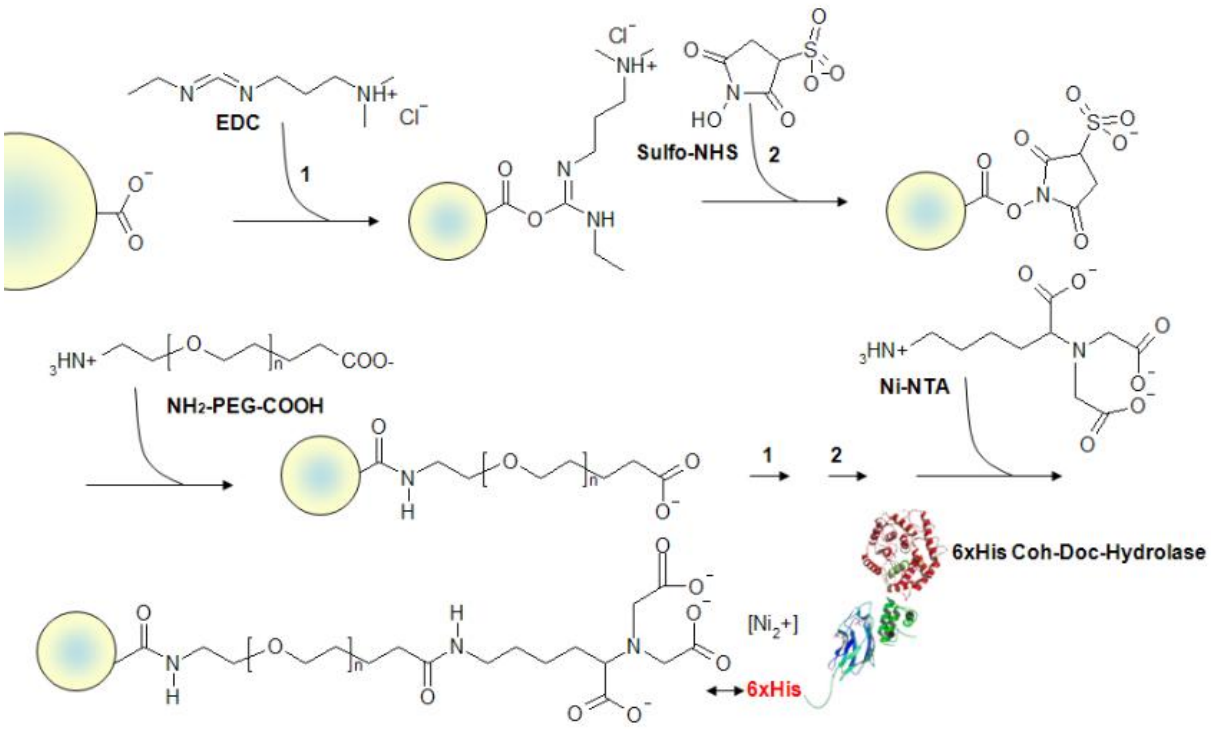


Fig. 14: Coupling steps with carboxylate modified nanoparticles. Surface modified particles are activated with EDC and sulfo-NHS. Active esters react with the free amino group of the hetero-bifunctional PEG based linker. By repetition of the activation steps the active carboxylate group of linker could react with the free amino group of NTA. After Ni^{2+} complexation by NTA, the His-tagged carrier proteins are immobilized on the surface of the particles via Histidin-nickel interaction.

3.5.2.2 NH_2-modified particles

Homo-bifunctional reagents specifically reacting with primary amine groups have been used extensively, as they are soluble in aqueous solvents and can form stable inter- and intra-subunit covalent bonds. Glutaraldehyde, a popular reagent, has been used in a variety of applications where maintenance of structural rigidity of a protein is important.

20 mg of amino modified nanoparticles were washed three times with 2 ml PBS buffer (10 mM NaH_2PO_4, pH 6.8) by separation with a strong disc magnet. The homo-bifunctional cross-linking agent glutaraldehyde was added to a final concentration of 2.5 % (v/v) and incubated for 5 hours at RT with slight shaking (see Fig. 15). Potentially formed Schiff-bases were reduced with 10 µl fresh $NaBH_4$ solution (2 mg/ml) for 5 hours at 4°C. 5 mg of O-(2-aminoethyl)-O-(2-carboxyethyl)-polyethylenglycol 3000 hydrochloride (NH_2-PEG-COOH) was added to the activated particles. The coupling reaction proceeded within 5 hours at RT. After incubation Schiff-bases were reduced and the following steps, starting with activation of carboxylate groups with EDC and sulfo-NHS, were performed as previously described (see 3.5.2.1).

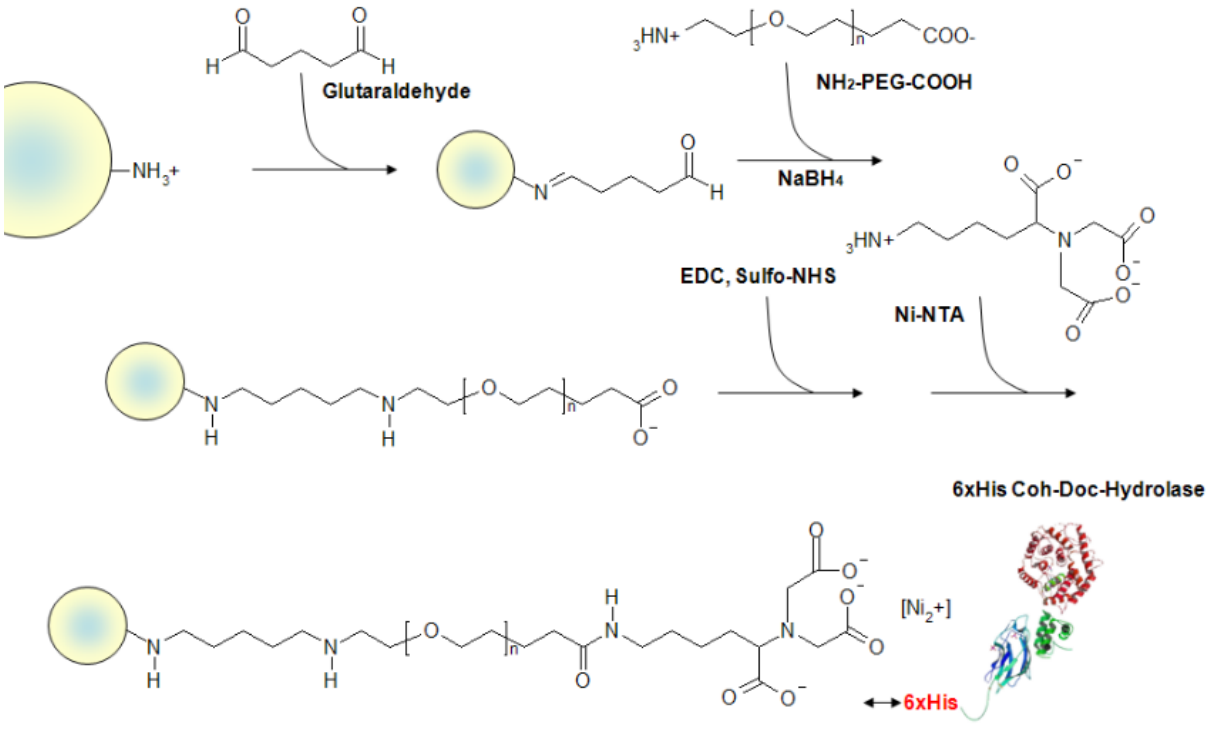


Fig. 15: Coupling steps with amino modified nanoparticles. Surface modified nanoparticles were activated with glutaraldehyde followed by interconnection of the PEG-based linker molecule and coupling of the enzymes analogous to the procedure with COOH modified beads (Fig. 14).

Chapter 4

Results

The technical task was to manufacture high-efficient enzyme complexes for the hydrolysis of crystalline cellulose by recombinant constructed scaffoldin proteins and hydrolases and the soluble components of *C. thermocellum* mutant, which lost the ability to produce the CipA protein.

4.1 Mutant of *Clostridium thermocellum*

In previous works by N. Schantz and V. Zverlov cultures of *C. thermocellum* were mutagenized and grown on cellobiose to enable the growth of cellulase defective mutants. Cells not adsorbing to crystalline cellulose were enriched by sequential steps of shaking a growing culture with crystalline cellulose and then transferring the non-adsorbed cells in the supernatant to a fresh culture vial. The last enrichment culture was diluted, plated on agar plates covered with a cellulose layer and incubated. Colonies with a reduced or absent ability to form clear halos in the cellulose around the colonies were picked and purified by single colony streaking under anaerobic conditions (Fig. 16) (Zverlov et al., 2005, Zverlov et al., 2008).

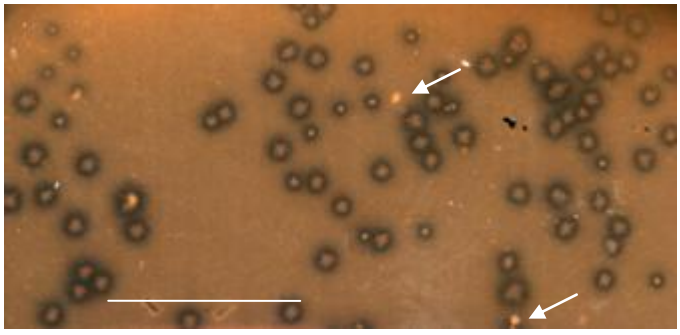


Fig. 16: Single colonies of a mutagenized culture of *C. thermocellum* enriched for non-adsorbing cells. Colonies on the turbid cellulose background produce a dark halo when degrading cellulose. The white bar indicates a length of 1 cm. Colonies with different sizes and abilities in cellulose hydrolysis can be recognized. Two colonies with reduced cellulose-degrading ability (putative mutants) are indicated by arrows.

Six of these "swimmer mutants", SM901 to SM906, were randomly selected. Molecular analysis confirmed that the mutants arose from the original *C. thermocellum* culture and were not infectants. One of the *C. thermocellum* mutants, SM901, had completely lost the ability to produce a scaffoldin protein or an active cohesin. The mutant produced the cellulosomal components in amounts approximately equal to the wild type, with the exception that the CipA component (the scaffoldin CipA) which was completely missing. The pattern of bands appearing in SDS-PAGE-separation of the proteins from purified cellulosomes was identical for WT and mutant culture supernatants except for the CipA band (Fig. 17).

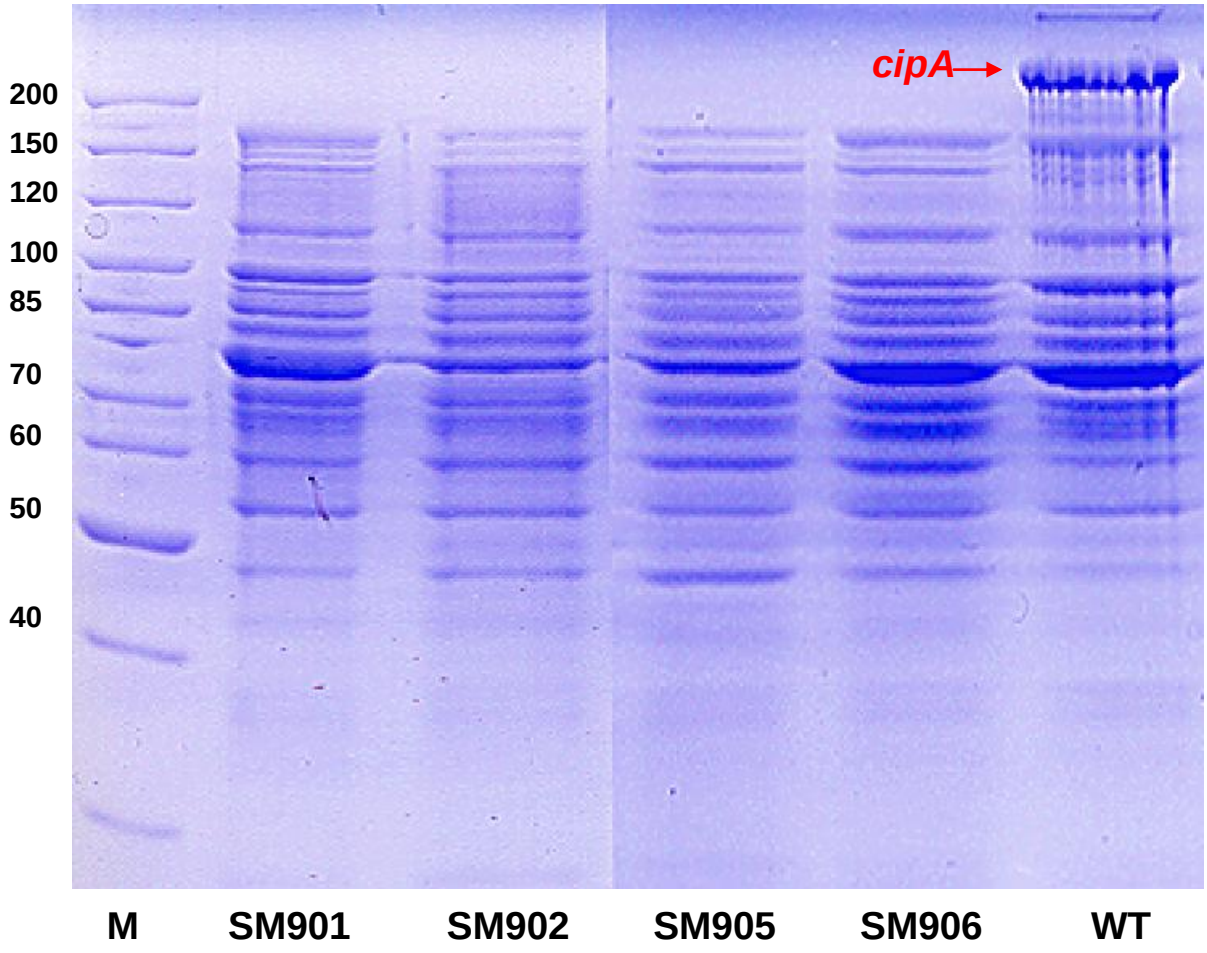


Fig. 17: SDS-PAGE of culture supernatant proteins from the mutants. Purified cellulosomes from the wild type (WT) were used. The position of the CipA protein is indicated. The left lane shows molecular mass markers (kDa)

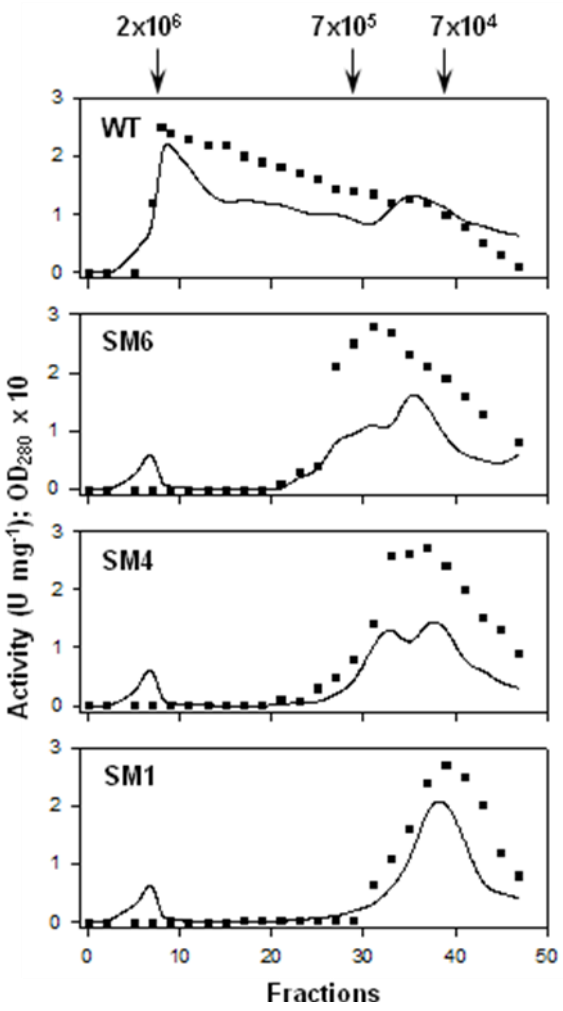


Fig. 18: Gel filtration (size exclusion chromatography) of culture supernatant proteins from the wild type (WT) and mutants. The amounts of protein (solid lines) and endoglucanase activity on barley β-glucan (dotted lines) are shown. The molecular masses of marker proteins (in kDa) are indicated above.

To analyze the presence of supramolecular complexes, the aktiv culture supernatants were subjected to size exclusion chromatography (Fig. 18). While most of the protein from the wild-type supernatant was found in high-molecular weight complexes (the cellulosomes), the sizes of these complexes were dramatically reduced in mutant SM906. The complexes were completely missing in mutant SM901. The lack of enzymatic complexes was even more obvious in the barley β-glucan activity profiles (Fig. 18). The proteins from the wild type contained in the peak slightly below 2 x 10^6 Da exhibited β-glucanase activity. Such high-molecular mass complexes were found to be completely absent in all mutants. The endoglucanase activity in mutants SM906 and SM904 showed the occurrence of small complexes that were not present in mutant SM901. However, such complexes were

significantly smaller than the cellulosomes. Mutant SM901 appeared to produce exclusively single enzyme components that appeared in a well-defined peak around the 70-kDa calibration marker (Zverlov et al., 2008).
Due to the lack of the scaffoldin protein CipA, it was evident that the *cipA* gene is involved in the mutant phenotype. Oligonucleotide primers were designed to investigate different regions of the *cipA* gene. Sequencing of the amplicons revealed the presence of a 1,447-bp IS element that was identical in all mutants. It was found to be inserted at four different locations within the *cipA* gene (Fig. 19). Two of the mutants obviously derive from identical mutation events. Part of the IS element showed homology with transposase genes of the IS*3* family. It was therefore called IS*1447*, referring to the length of the sequence. The transposase gene was transcribed in all cases in a direction opposite to the *cipA* gene. The inverted repeats were 26 bases in length and were situated at the far ends of the IS element. The insertion at four different locations in *cipA* of one copy of IS*1447* allowed the detection of a 3-base pair target sequence that was different in each insertion event. This non-specific target region was duplicated upon insertion.

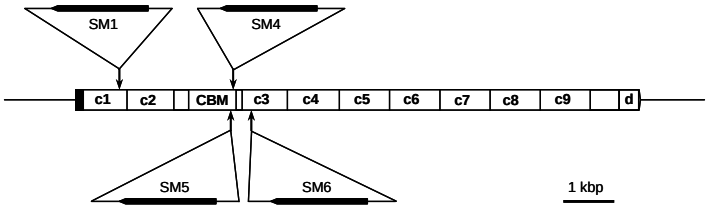


Fig. 19: Structure of the *cipA* gene and positions of IS*1447* insertions (mutant designations are indicated). c1, cohesin module 1; CBM, carbohydrate binding module; d, dockerin module.

Concentrated cell-free culture supernatants of the wild type (having cellulosomes) and mutant SM901 (free enzymes without organization in a cellulosome) were subjected to enzymatic tests on barley ß-glucan, CMC (both control), and micro-crystalline cellulose MN300 and Avicel. The enzymatic activity on barley ß-glucan and CMC were about 8.0 and 1.0 U mg^{-1} protein respectively. In contrast, specific activity on crystalline cellulose was dramatically reduced in the mutant SM901, up to 15 fold compared to the wild type (Tab. 4).

Substrate	SM901 mutant [U/mg]	Wildtype [U/mg]
β-glucan	7.9 ± 1.1	9.5 ± 0.9
CMC	1.1 ± 0.1	1.2 ± 0.1
MN300	0.03 ± 0.01	0.42 ± 0.11
Avicel	16.5 ± 1.4 x 10^{-3}	190 ± 5.5 x 10^{-3}

Tab. 4: Enzymatic activity of concentrated culture supernatants of the mutant SM901 and the wildtype on 0.5% barley ß-glucan, CMC (soluble), MN300 cellulose, Avicel (crystalline).

Functionally intact cohesins in the residual, truncated CipA proteins were detected on SDS-PAGE slabs by affinity blotting with the recombinant dockerin containing marker protein rCel9N and anti-His tag antibodies after *in situ* renaturation. Only proteins were detected specifically interacting with rCel9N protein via cohesin-dockerin binding (Fig. 20).

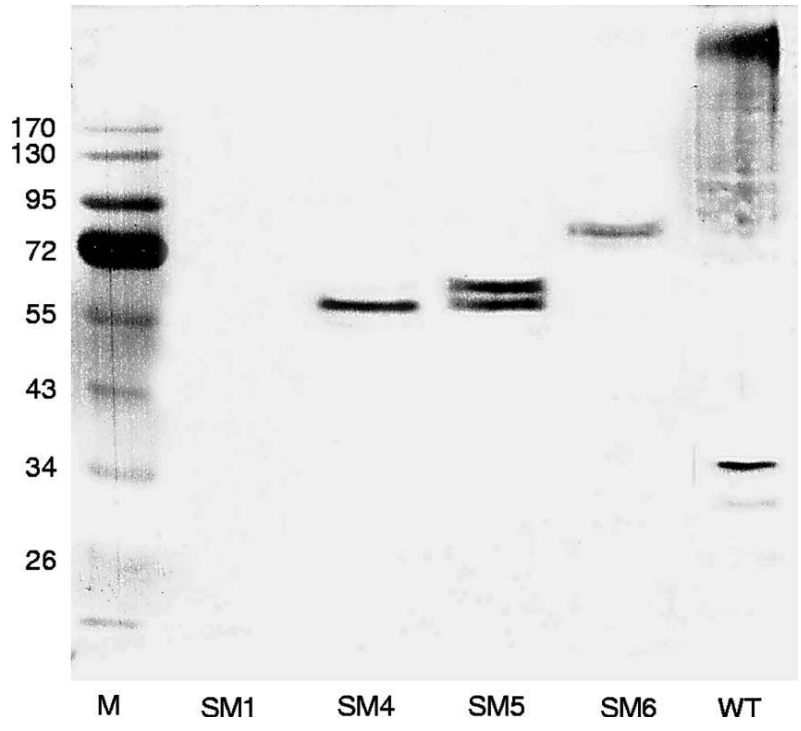


Fig. 20: Denaturing gel electrophoresis of concentrated culture supernatants. CipA fragments containing cohesins. Proteins containing cohesin modules were detected with His tag-Cel9N and anti-His antibodies. The numbers indicate molecular masses (kDa) of the protein mass standard (lane M).

4.2 Miniscaffoldins and CipA

The mutant SM901 thus produced a non-complexed mixture of the native cellulosomal proteins without the complex integrating protein CipA the scaffoldin. These mutant supernatant proteins were used to reconstitute an artificial cellulosome and to measure the effect of complexation, either with native enzymes or with enzymes isolated from recombinant hosts. For immobilization on nanoparticles, recombinant mini-scaffoldins consisting of various cohesins with or without a carbohydrate binding module were produced. Sample structures are depicted in Fig. 21.

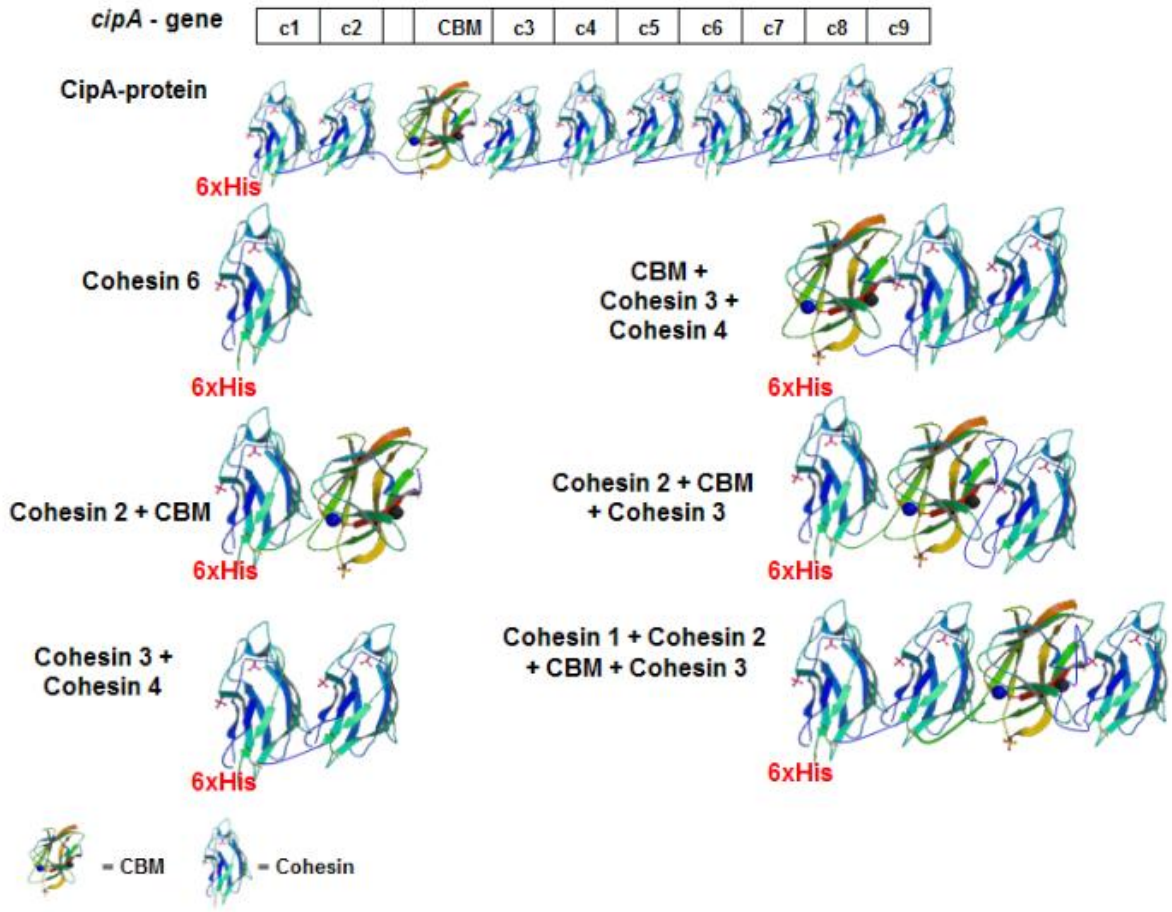


Fig. 21: Schema of the recombinant scaffoldin constructs. CipA-protein and derivatives of the CipA-protein of *C. thermocellum* (top row: c1 = cohesin 1 etc.; CBM = carbohydrate binding module). 3D-structures (curtesy H. Gilbert) are added for illustration.

Six different miniscaffoldins were designed, each containing different numbers of cohesins. Miniscaffoldins are based on the cellulosomal scaffoldin from *C. thermocellum*, in which two of the nine cohesins (Cohesin 2 and cohesin 3) are separated by an internal CBM. Single cohesins, which

lack a CBM, were designed in order to determine whether cohesin-dockerin interaction between miniscaffoldins and SM901 mutant exoenzymes is specific and whether simple complexation of enzymes would also promote synergism.

The miniscaffoldins are N-terminally fused with a 6xHis-tag for purification via FPLC and immobilization on nanoparticles. Figure 22 shows SDS-PAGE of purified miniscaffoldins.

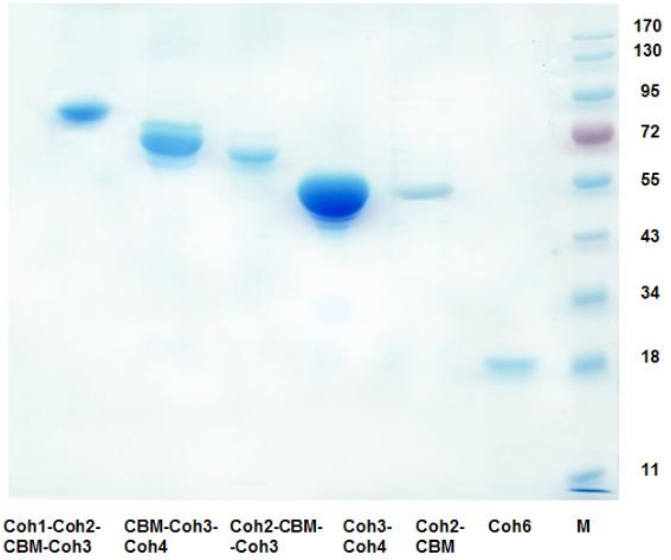


Fig. 22: SDS-PAGE of purified recombinant miniscaffoldins. Coh = cohesin, CBM = carbohydrate binding module; the right lane shows molecular mass markers (kDa)

Specific interaction between miniscaffoldins and mutant exoenzymes was determined by FPLC and additional SDS-PAGE analysis. Sequence comparisons (Fig. 23) of the nine scaffoldin internal cohesins showed that cohesin 1 and 2 differ most from the other cohesins. However, sequences of remaining cohesins only vary in single bases. Therefore the extreme cohesins 1 and 6 were chosen to determine specific interaction abilities with dockerin bearing hydrolases from mutant supernatant. A single cohesin (1 or 6 respectively) was immobilized via Ni^{2+}-Histidin-interaction on a Histrap-column and mutant supernatant with exoenzymes was rinsed over the column. Dockerin bearing hydrolases bound to immobilized cohesins. Eluated fraction, flow through and exo-proteins were applied to SDS-PAGE and the elution profile of the two different cohesins was evaluated.

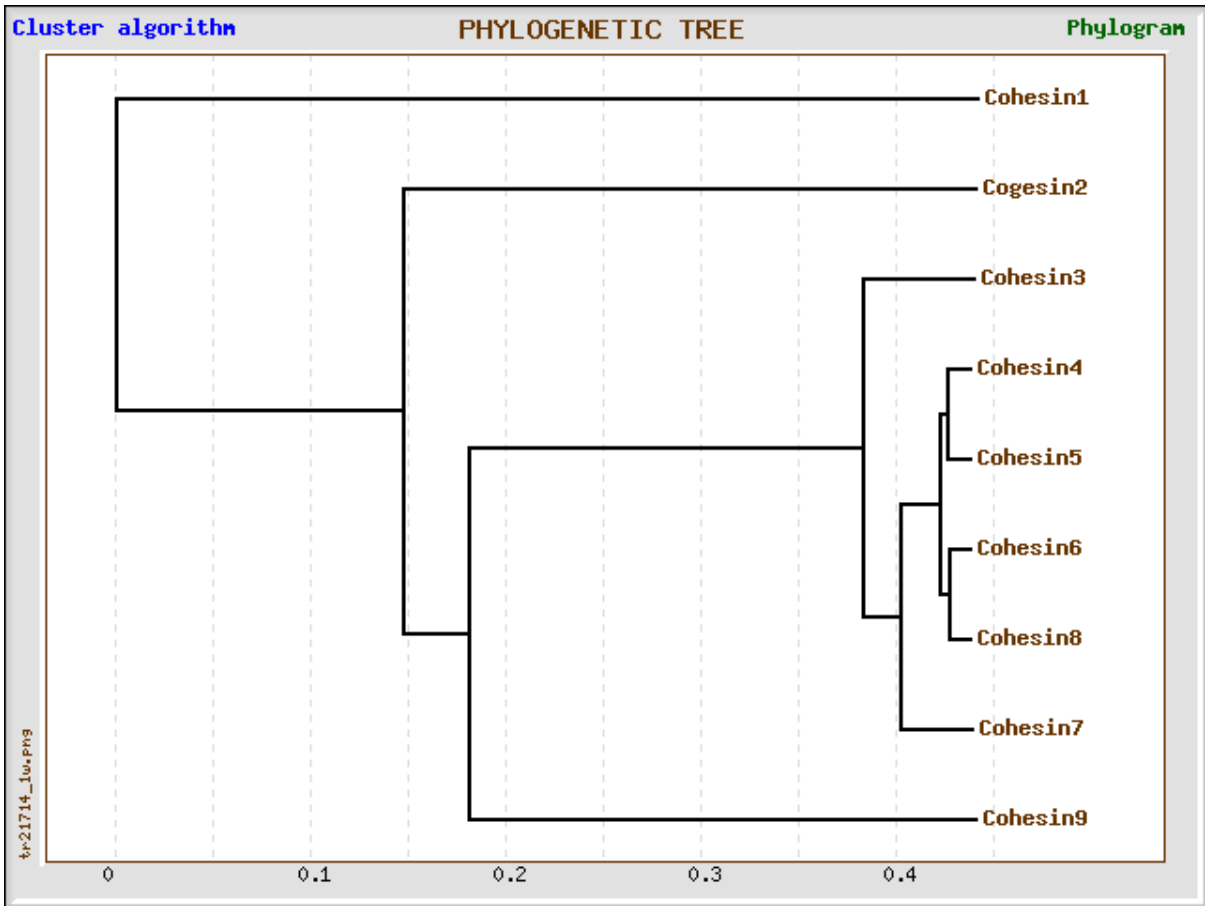


Fig. 23: Phylogenetic tree of the nine cohesins of *Clostridium thermocellum*. Cohesins 3 to 9 show high sequences similarities, whereas sequences of cohesin 1 and 2 differ from remaining cohesins.

The SDS-PAGE clearly demonstrated the unspecific binding abilities of different cohesins with dockerin bearing hydrolases (Fig. 24). The pattern of cohesin 1 which interacted with mutant hydrolases showed no significant differences to the pattern of hydrolases bound to cohesin 6. In both cases, analysis of the flow through showed that a high amount of proteins passed the cohesins, indicating that binding modules of the applied cohesins were satisfied with SM901 proteins. However, the mutant supernatant contained some cellulosomal structure proteins presumably, such as SdbA, OlpB and S-layer proteins which do not interact with cohesins.

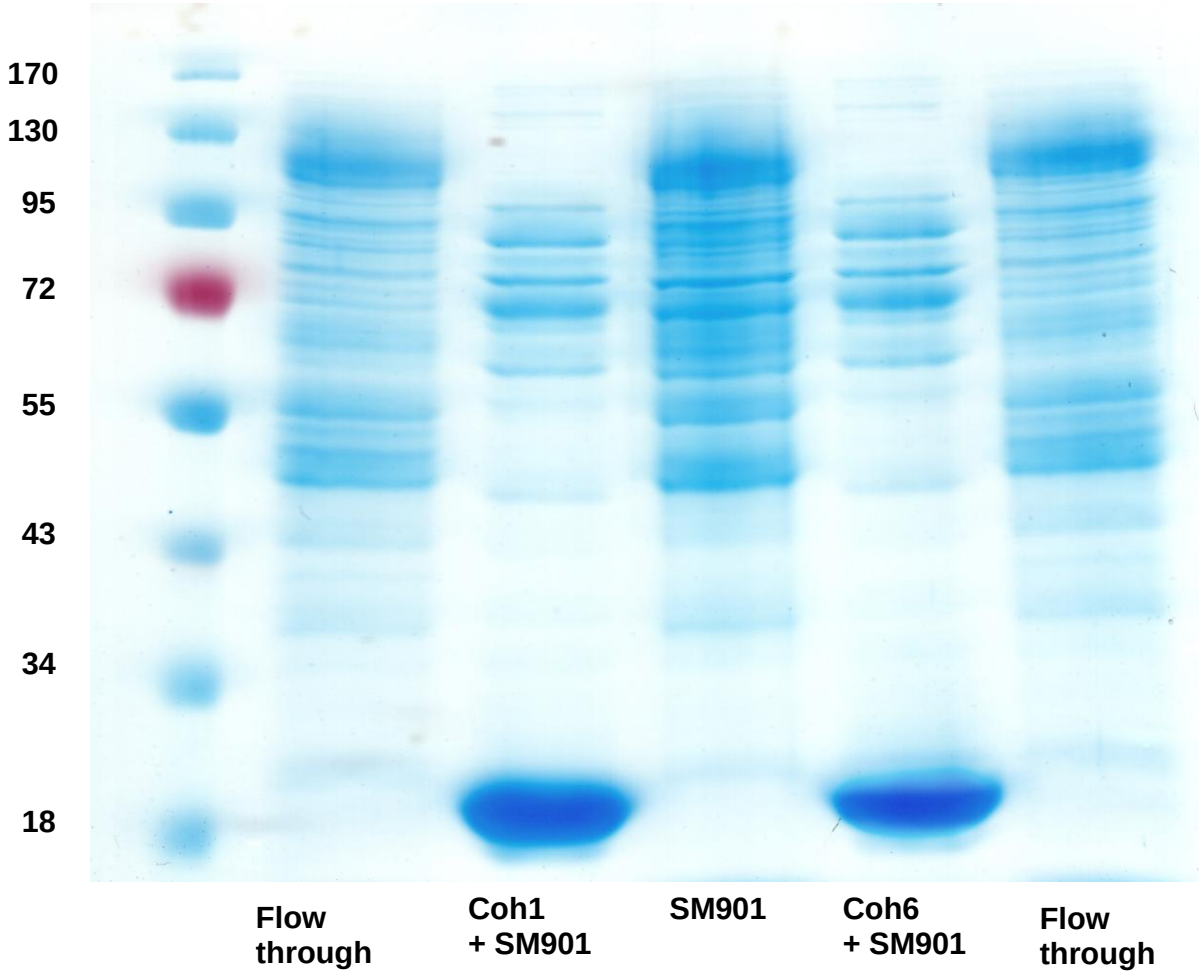


Fig. 24: Elution profile of two different cohesins interacted with SM1 exoenzymes. Cohesin 1 and 6 were applied to His-trap column and rinsed with SM901 supernatant; lane 1: flow through; lane 2: cohesin 1 + SM901 proteins; lane 3: SM901 exo-proteins, lane 4: cohesin 6 + SM901 proteins, lane 5: flow through; left lane shows molecular mass markers (kDa)

Complex formation in the presence of calcium was verified using non-denaturing PAGE. Nondenaturing PAGE clearly demonstrated that near complete complex formation could be achieved simply by mixing the desired components *in vitro*. An example is shown in figure 25. Cellobiohydrolase Cel9K was mixed with CBM-cohesin3-cohesin4 and complex formation occurred with miniscaffoldin. The stoichiometric mixtures of the enzyme and miniscaffoldin resulted in a single band with altered mobility thus indicating that complete or near complete complexation was achieved.

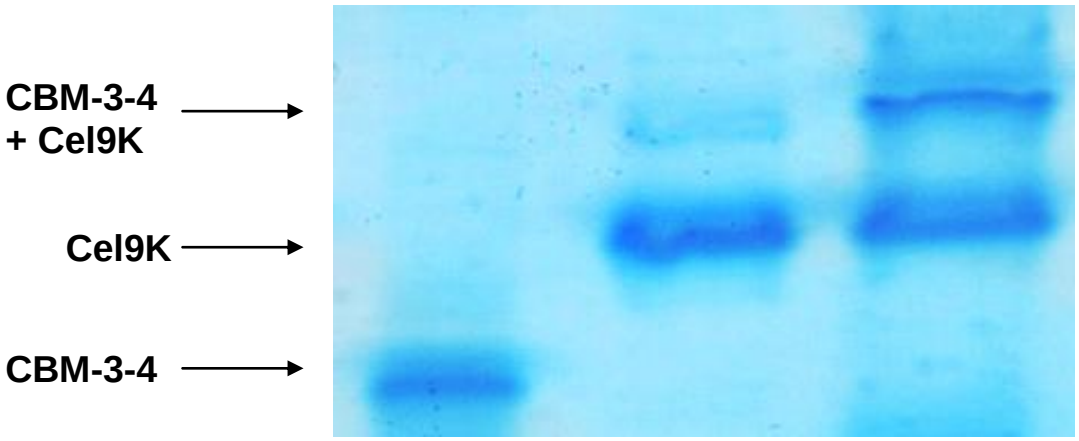


Fig. 25: Assembled complex on nondenaturing gel. First two lanes show the single components CBM-cohesin3-cohesin4 and cellobiohydrolase Cel9K, respectively. Lane 3 shows complex formation between the single components. In each lane, equimolar concentrations of the indicated proteins were used. Similar quality gels were obtained for other miniscaffoldins.

Preliminary tests showed that a sufficient expression of the complete CipA protein was not possible in *E. coli* conventionally. Due to the low G+C content in the genome of *Clostridium thermocellum* codons are preferred which contain adenine or uracil. With the heterologous expression of genes of *C. thermocellum* in *E. coli*, rare codons could act as a translation barrier (Young et al., 1989). A comparison in *E. coli* rarely used codons with their appropriate frequency in clostridial genes showed significant differences (see Tab. 5).

Organism/Protein	Arginine (AGG/AGA)	Leucine CTA	Isoleucine ATA
E. coli	0.14%/0.21%	0.32%	0.41%
C. thermocellum	0.51%/2.08%	0.71%	3.91%
CipA	0.37%/1.67%	0%	6.08%

Tab. 5: Codon-usage in *E. Coli*, *C. Thermocellum*, CipA protein. Values show the relative frequency of rare codons in bacterial genes in [%].

A T7 polymerase driven pET101 vector containing CipA gene was co-expressed with pLysRARE plasmids in Rosetta-gami B (DE3)pLysS cells. This co-expression vector contains genes for different tRNAs (*argU*, *ileY*, and *leuW*). This system allowed achievement of *cipA* gene product (see Fig. 26).

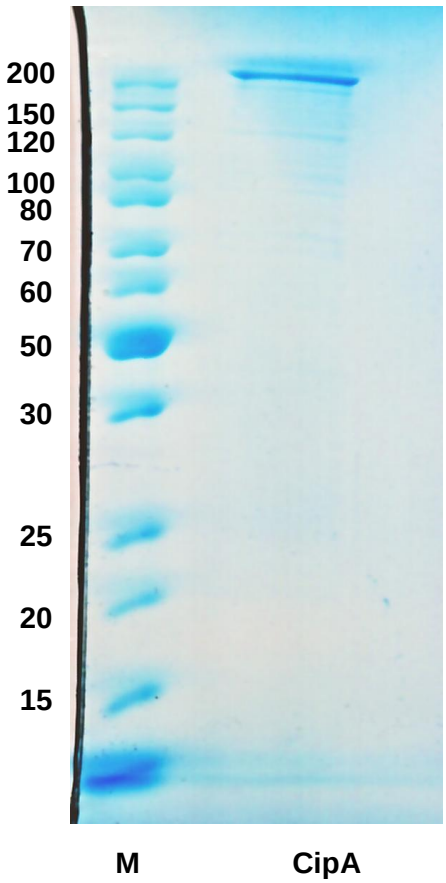
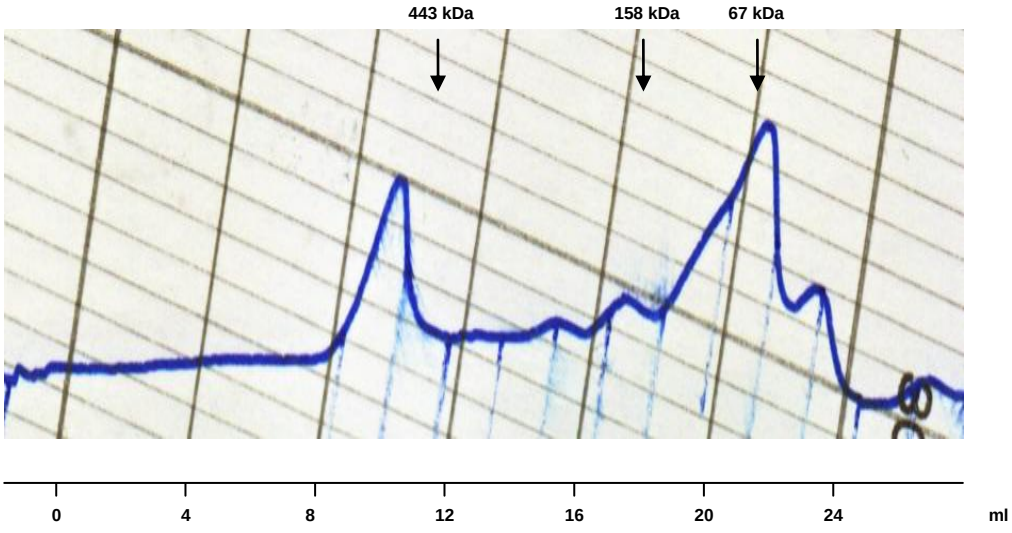


Fig. 26: SDS-PAGE of purified rekombinant CipA. The left lane shows molecular mass markers (kDa).

CipA protein was incubated with SM901 proteins in the presence of 5 mM $CaCl_2$ and size exclusion chromatography was performed (Fig. 27). According to the standards, a peak appeared at a molecular mass higher than 443 kDa, indicating complex formation of 200 kDa CipA protein with at least 7 SM901 mutant dockerin bearing enzymes occurred. Collected fractions of this peak showed hydrolytic activity towards β-glucan comparable to low-molecular fractions.

Fig. 27: Gel filtration (size exclusion chromatography) of CipA protein incubated with SM901 mutant enzymes. Black arrows indicate molecular mass of used standards. Apoferritin (443 kDa), Aldolase (158 kDa), Bovine serum albumin (67 kDa).



Miniscaffoldins and CipA proteins were used to determine potentially occurring synergistic effects of neighboring enzymes. Additionally miniscaffoldins either with or without CBM were also investigated to verify the impact of CBM on cellulolytic activity. The SM901 mutant exo-enzymes were mixed stoichiometrically with single miniscaffoldins or CipA. After incubation with different soluble and insoluble substrates at 60 °C for 10 to 30 minutes for soluble substrate and 24 hours for insoluble substrate the amount of produced reducing ends were determined by a DNSA-test. The enzymatic activities on barley β-glucan and CMC were about 8.0 and 1.0 U/mg protein respectively in all tested systems (Tab. 6). In contrast, miniscaffoldins containing two or more cohesins were generally found to be more active on insoluble substrates than simple mixtures of the free enzymes.

Substrate	free Enzymes	c1	c2-CBM	c1-c2	c2-CBM-c3	CBM-c3-c4	c1-c2-CBM-c3	CipA	Cell.
β-glucan	7912	7634	8014	8045	7845	7792	8056	8178	8267
CMC	1134	1065	1129	1178	1067	1272	1108	1201	1267
PASC	2731	2847	2785	2629	2756	2683	2741	2850	2782
MN300	30	32	67	63	112	115	145	373	423
Avicel	12	13	25	19	39	40	57	150	190

Table 6: Enzymatic activity [mU/mg protein] of free enzymes free *Clostridium thermocellum* SM901 mutant enzymes, incubated with different miniscaffoldins and recombinant CipA protein on 0.5% barley ß-glucan, CMC (soluble), PASC (amorphous), MN300 cellulose, Avicel (crystalline), and for comparison spec. activity of native purified *Clostridium thermocellum* cellulosome. c, cohesin; CBM, carbohydrate binding module; Cell., cellulosome. Each value is the average of triplicate measurements.

Bringing two or more cellulolytic modules into close proximity clearly enhances the catalytic efficiency on crystalline cellulose. The specific activity was consistently increased with increasing number of cohesins in miniscaffoldin compared to the degradation rates of free enzymes. The enzymatic activities of free enzymes on MN300 and Avicel were about 30 and 12 mU/mg protein respectively. However, the values of enzymes bound to miniscaffoldin containing three cohesins or to recombinant CipA were about 5 (c1-c2-CBM-c3) and 13 (CipA) times higher respectively. The observed activity climbed with incubation time and reached a maximum after 24

hours. The cellulolytic activity of the enzyme complexes was further improved in scaffoldins that contained a CBM. Cohesins connected with a CBM doubled catalytic activity on crystalline cellulose compared to just a single cohesin without CBM.

4.3. Effectors and inhibitors

The effects of various added chemicals on the SM901 mutant enzymes were determined and summarized in table 7. The enzymes were affected by bivalent cations with the exception of $ZnCl_2$. The addition of 10 mM $CaCl_2$ to the mixture resulted in an enhancement in specific activity of 19.3 mU/mg to 25.7 mU/mg. However, degradation rates were significantly diminished after treatment with other cations such as Fe^{2+}, Hg^{2+} and Cu^{2+}. Furthermore, diminished or no activities could be observed after the addition of DTNB, SDS, EDTA, DTT and Ethanol. No significant changes occurred with Tween and Triton X 100. The addition of DMSO has a positive effect on hydrolysis, whereas a drastic increase in hydrolytic activity of about 250 % was determined after BSA was added to the enzymes mixture.

Substance	Concentration	Specific activity [mU/mg]
None	----	19.3 ± 2.1
$CaCl_2$	10 mM	25.7 ± 2.5
$CuSO_4$	10 mM	3.1 ± 0.5
$FeCl_3$	10 mM	7.8 ± 1.2
$HgCl_2$	10 mM	4.8 ± 1.3
$ZnCl_2$	10 mM	17.5 ± 2.0
DTNB	1 mM	0.0 ± 0.0
SDS	10 mM	7.3 ± 2.4
EDTA	10 mM	13.9 ± 2.0
DMSO	10%	25.5 ± 2.9
DTT	1mM	5.4 ± 0.8
Tween	2 mM	18.1 ± 1.9
Triton X 100	2 mM	17.8 ± 2.1
BSA	10 mg/ml	49.7 ± 3.4
Ethanol	10%	12.3 ± 2.5

Tab. 7: Effect of chemicals and inhibitors on enzyme activity of SM901 mutant proteins. Specific activities [mU/mg] were determined on 0.5 % Avicel.

4.4 Key components for a cellulase complex

In previous studies of Gold et al. (2007) metabolic isotope-labelling strategies were used in conjunction with nano-liquid chromatography-electrospray ionization mass spectrometry peptide sequencing to assess quantitative alterations in the expression patterns of subunits within cellulosomes of the cellulolytic bacterium *Clostridium thermocellum*, grown on either cellulose or cellobiose. Proteins that exhibited higher expression in cellulosomes from cellulose-grown cells than in cellobiose-grown cells were the cell surface anchor protein OlpB, exoglucanases Cel48S and Cel9K, and the glycoside hydrolase family 9 (GH9) endoglucanase Cel9J. Conversely, lower expression in cellulosomes from cells grown on cellulose than on cellobiose was observed for the GH8 endoglucanase CelA; GH5 endoglucanases CelB, CelE, CelG; and hemicellulases XynA, XynC, XynZ, and XghA. GH9 cellulases were the most abundant group of enzymes per CipA when cells were grown on cellulose, while hemicellulases were the most abundant group on cellobiose (Gold et al., 2007). Similar results are described in previous studies of Zverlov et al. (2005): The 13 major components in the cellulosome were identified and their approximate abundance in the cellulosomal particles was calculated (Zverlov et al., 2005). This suggests the main proteins that might be necessary and sufficient to reconstruct an efficient *in vitro* cellulase system.

Three exoglucanases Cbh9A, Cel9K and Cel48S, as well as two endoglucanases, Cel9J and Cel9R, were produced recombinantly and added in different fractions to the SM901 mutant supernatant to increase hydrolytic activity on crystalline cellulose. Additionally some xylanases - Xyn11A, Xyn10C, Xyn10Z and Xgh74A - were cloned and expressed in order to investigate hydrolytic behaviour of complexed xylanases on insoluble components of oat spelt xylan. In following sections the single recombinant components are characterized.

4.4.1 Cellobiohydrolase Cbh9A, exoglucanase Cel9K, exoglucanase Cel48S

The cellobiohydrolase Cbh9A from *C. thermocellum* exhibits a multidomain structure (Fig. 28, A) of unusual complexity. It consists of an N-terminal cellulose binding module (CBM) homologous to CBM family IV, an immunoglobulin like β-barrel module, a catalytic module homologous to cellulase family E1, a duplicated module similar to fibronectin type III (Fn3) modules, a CBM homologous to family III, a highly acidic linker region, and a C-terminal dockerin module (Zverlov et al., 1998).

The exoglucanase Cel9K of *C. thermocellum* has module structures consisting of a signal peptide, a family IV cellulose binding module, a family 9 glycosyl hydrolase module, and a dockerin module (Fig. 28, B). Cel9K shows some similarities to Cbh9A. A distinguishing feature between the two polypeptides is that there is a 330-amino-acid insertion in Cbh9A between the catalytic module and the dockerin module containing a fibronectin type 3-like domain and family III CBM (Zverlov et al., 1998).

The glucanase Cel48S, the most abundant catalytic subunit of the *Clostridium thermocellum* cellulosome, displayed exoglucanase characteristics (Wang et al., 1993, Wang et al., 1994). It consists of a family 48 glycosyl hydrolase module and a C-terminal dockerin module (Fig. 28, C).

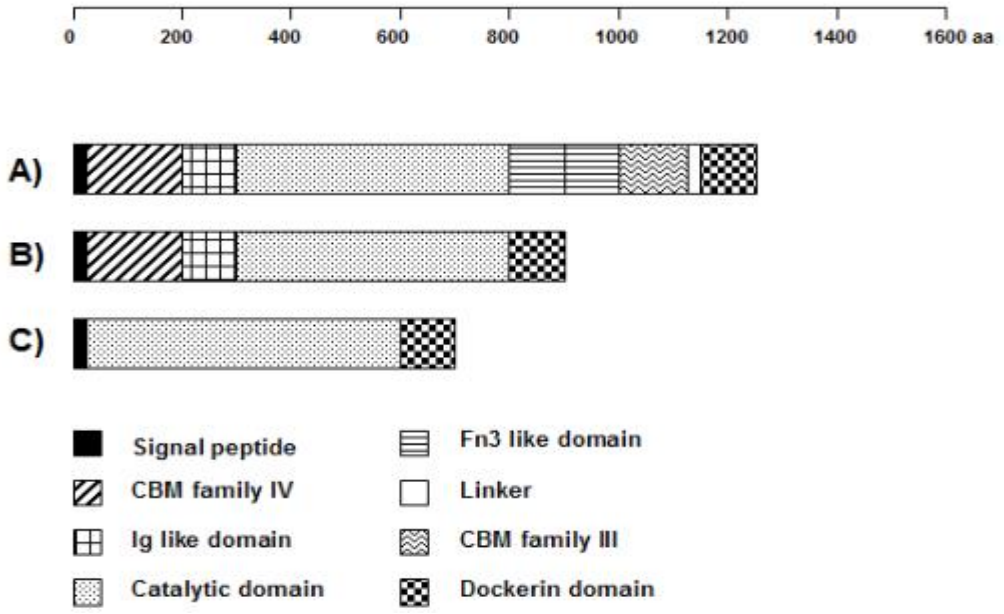


Fig. 28: Domain structure of native exoglucanases. A) Cellobiohydrolase Cbh9A. **B)** Cellobiohydrolase Cel9K. **C)** Exoglucanase Cel48S

Recombinant pQE (30-32) vectors carry the nucleotide sequence of one of the three different exoglucanases with removed signal sequence. The native protein Cbh9A has a molecular mass of ~138 kDa. However, SDS-PAGE analysis of purified rekombinant Cbh9A shows a single band with molecular mass of ~100 kDa (Fig. 29). SDS analysis from Koeck (2009) showed similar electrophoretic mobility of Cbh9A. The CbhA9A protein could be verified by MALDI-TOF. The gene *cel9K* has an open reading frame of 2,685 nucleotides coding for a polypeptide of 895 amino acid residues with a calculated mass of ~101 kDa. SDS-PAGE of recombinant Cel9K shows a single band with molecular mass of ~ 100 kDa as expected (Fig. 29), indicating that glycosylation of the polypeptide was negligible, if present at all.

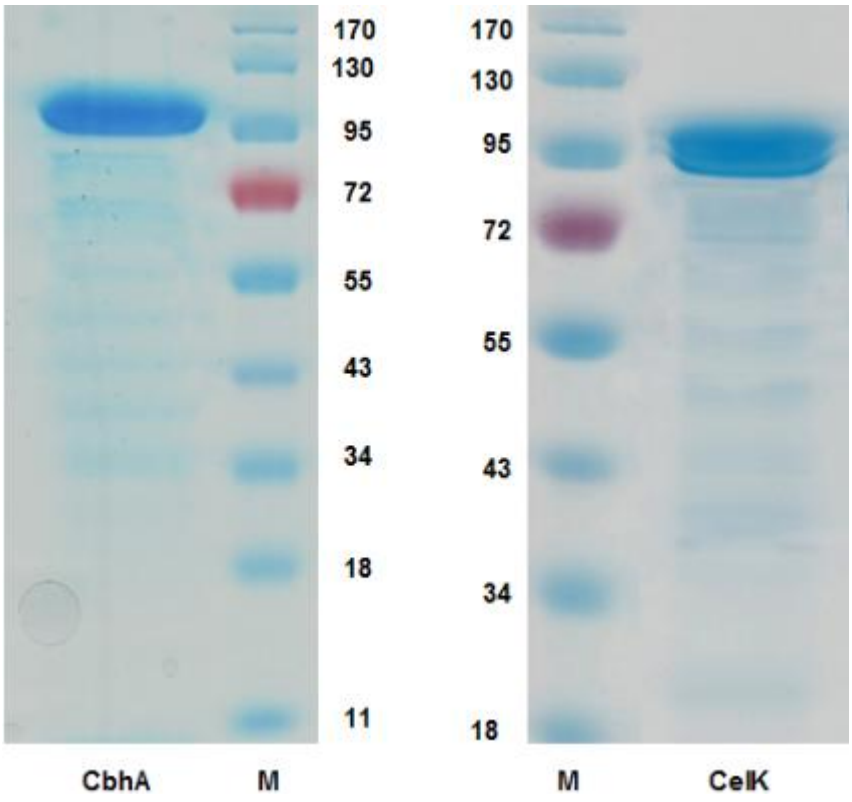


Fig. 29: SDS-PAGE of purified recombinant Cbh9A (left) and Cel9K (right). Lane M shows molecular mass markers (kDa).

Subsequent expression of the cloned gene encoding Cel48S resulted in a protein present as inclusion bodies. The accumulated proteins were purified by denaturation with urea and additional denaturation by dialysis. However, activity assays on insoluble cellulose showed dramatically diminished

specific activity compared to literature sources (Kruus et al., 1995). Purified recombinant Cel48S has a molecular mass of ~ 85 kDa as expected.

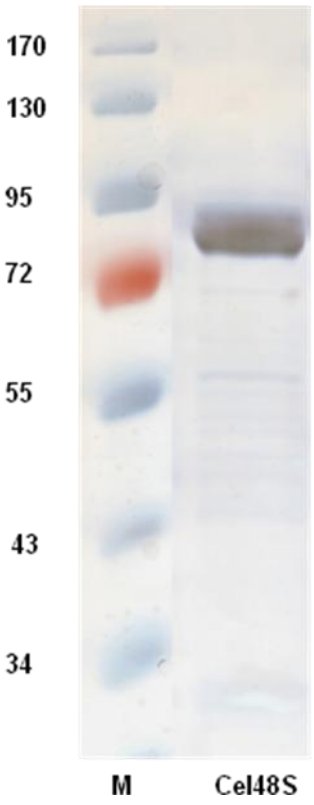


Fig. 30: SDS_PAGE of purified exoglucanase Cel48S after dialysis. Left lane shows molecular mass marker in (kDa).

Activity assays on different soluble and insoluble substrates were carried out. Table 8 shows the specific activity [mU/mg protein] of the different recombinant exoglucanases. The enzymes exhibit a specific activity of about 8,000 mU/mg protein on barley β-glucan, but show no significant activity on carboxy-methyl cellulose (CMC). Cbh9A, Cel9K and Cel48S show exoglucanase activity and hydrolysis rates are retarded by outstanding carboxy-methyl groups of substrate. On amorphous substrate (PASC) the enzymatic activities were ~1630 ± 220 mU/mg protein. The degradation of the insoluble, cristallinic substrates MN300 and Avicel was diminished to values of 15.9 (Cbh9A), 17.8 (Cel9K), 22.7 (Cels48S) and 4.2 (Cbh9A), 5.3 (Cel9K), 8.5 (Cel48S) mU/mg protein, respectively. These results are in

accordance with previous studies, albeit not all of these substrates used in this study were tested (Kataeva et al., 2002, 1999; Wang et al. 1994).

Substrate	Cbh9A	Cel9K	Cel48S
		[mU/mg]	
β-glucan	8290 ± 541	8102 ± 579	7832 ± 754
CMC	n.d.	n.d.	< 1.0
PASC	1661 ± 200	1623 ± 247	1598 ± 261
MN300	15.9 ± 2.8	17.8 ± 3.5	22.7 ± 3.7
Avicel	4.2 ± 0.9	5.3 ± 1.2	8.5 ± 2.1

Tab. 8: Hydrolytic activity [mU/mg protein] of rekombinant cellobiohydrolase Cbh9A, cellobiohydrolse Cel9K and exoglucanase Cel48S on 0.5% soluble (barley β-glucan, CMC) amorphous (PASC) and insoluble (MN300, Avicel) cellulose. n.d., not detectable.

4.4.2 Endoglucanase Cel9J, Cel9R

Cel9J is a modular enzyme composed of N-terminal signal peptide and six modules in the following order: an S-layer homology module, a cellulose binding module (CBM) homologous to CBM family 30, a subfamily E1 endoglucanase module, a family J endoglucanase module, a docking module, and a module of unknown function (Fig. 31, A) (Ahsan et al., 1996).

Cel9R, a major component in the cellulosome of *Clostridium thermocellum*, is one of the most prevalent β-glucanases in the complex after Cel48S and Cel8A. The native enzyme consists of an endoglucanase module, a cellulose binding module homologous to CBM family III and a C-terminal dockerin module (Fig. 31, B) (Zverlov et al., 2005).

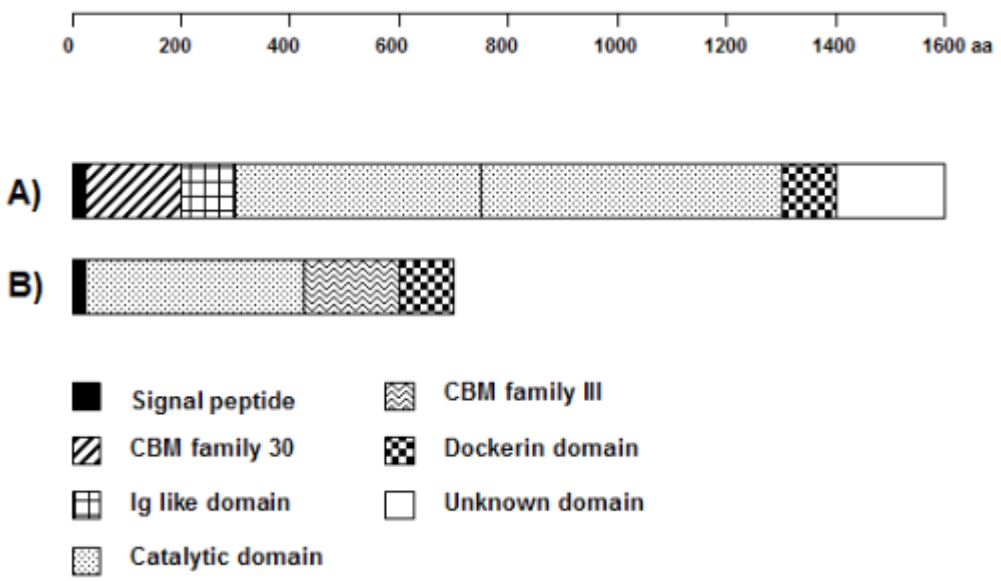


Fig. 31: Domain structure of native endoglucanase. A) Endoglucanase Cel9J **B)** Endoglucanase Cel9R

Recombinant pQE-vectors carry the nucleotide sequence of one of the two different endoglucanases with removed signal sequence. The native protein Cel9J has a molecular mass of ~178 kDa. The correct size of recombinant Cel9J could be confirmed by SDS-PAGE analysis of the purified enzyme (Fig. 32). The gene *cel9R* has an open reading frame of 2,706 nucleotides coding for a polypeptide of 903 amino acid residues with a calculated mass of ~85 kDa. SDS-PAGE of recombinant Cel9R shows a single band with a molecular mass of ~ 85 kDa as expected (Fig. 32).

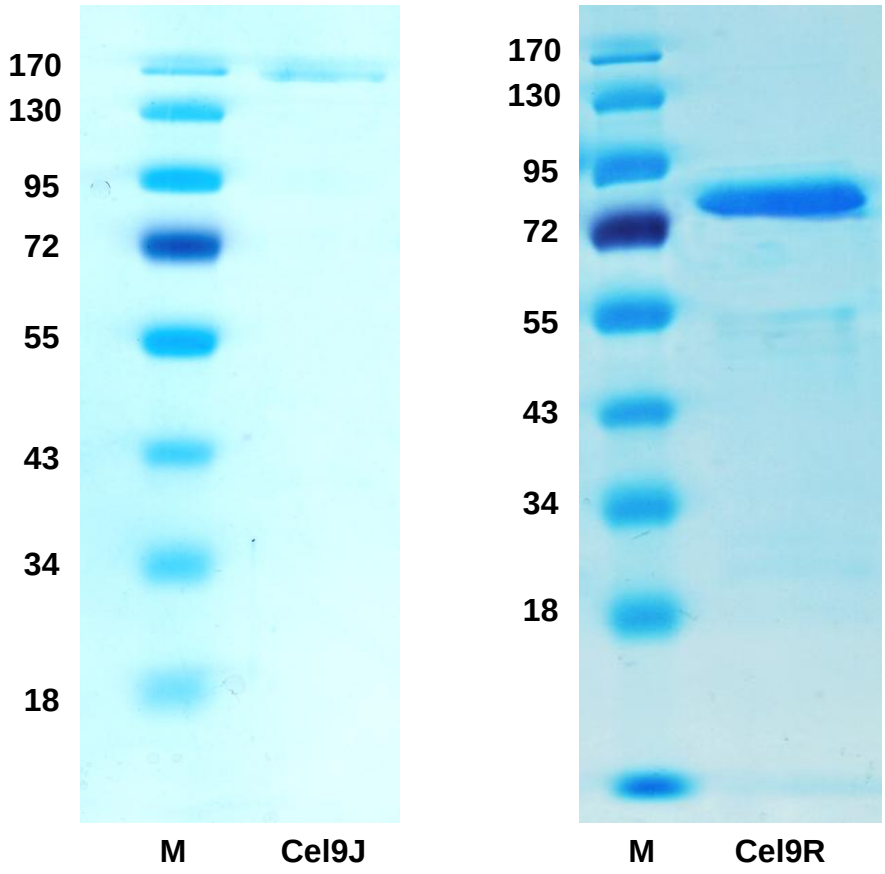


Fig. 32: SDS-PAGE of purified recombinant Cel9J (left) and Cel9R (right). The left lane shows molecular mass markers (kDa).

Activity assays on different soluble and insoluble substrates were carried out. Table 9 shows the specific activity [mU/mg protein] of the different recombinant endoglucanases.

Substrate	Cel9J	Cel9R
	[mU/mg]	
β-glucan	16430 ± 1386	27245 ± 3048
CMC	6586 ± 842	8655 ± 1068
PASC	2235 ± 363	2534 ± 298
MN300	24.6 ± 2.7	23.4 ± 1.9
Avicel	8.3 ± 1.1	7.3 ± 1.8

Tab. 9: Hydrolytic activity [mU/mg protein] of rekombinant endoglucanase Cel9J and endoglucanase Cel9R on 0.5% soluble (barley β-glucan, CMC) amorphous (PASC) and insoluble (MN300, Avicel) cellulose.

The enzymes exhibited strong hydrolytic activities on barley β-glucan. Due to their endoglucanase mode of action the hydrolases were able to degrade CMC. On amorphous substrate (PASC) the enzymatic activities were about

2.2 ± 3.6 (Cel9J) and 2.5 ± 2.9 (Cel9R) U/mg protein for Cel9J and Cel9R respectively. The degradation rates of the insoluble, cristallinic substrates MN300 and Avicel were diminished to values of 24.6 (Cel9J), 23.5 (Cel9R) and 8.3 (Cel9J), 7.3 (Cel9R) mU/mg protein.

4.4.3 β-Glucosidase BglB from *Thermotoga neapolitana* fused with dockerin module from endoglucanase CelA

To gather more information on the hydrolytic mode of action by hydrolysis of soluble and insoluble substrates, the degradation products were analysed with thin-layer chromatography (TLC). Degradation of soluble cellulose (barley β-glucan) by SM901 mutant enzymes as well as by native cellulosomes resulted mainly in disaccharides, cellobiose, and oligosaccharides like cellotriose and cellotetrose, rather than glucose. Disruption of insoluble cellulose (Avicel) yielded mainly cellobiose with marginal quantities of glucose (Fig. 33).

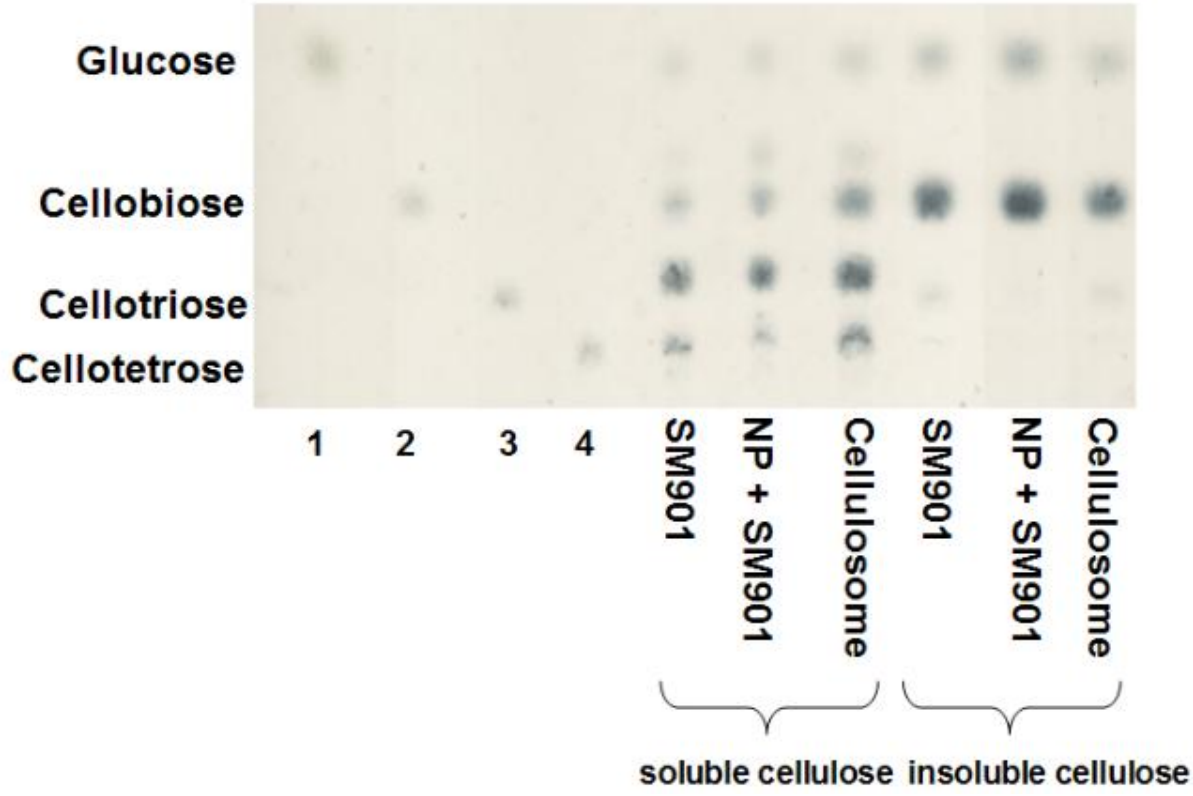


Fig. 33: Thin-layer chromatography of hydrolytic products of degradation of insoluble and soluble cellulose by SM901 mutant hydrolytic enzymes, SM901 supernatant proteins immobilized on nanoparticles and native cellulosome. Lane 1 to 4 show marker from glucose to cellotetrose. Lane 5 to 7 and 8 to 10 shows degradation products from mutant enzymes, immobilized on nanoparticles and cellulosome on soluble substrate (barley β-glucan) and insoluble substrate (Avicel), respectively.

A problem encountered with cellulase-mediated saccharification is the inhibition of cellulases by cellobiose, the major immediate product of cellulase action. A strong product inhibition is a major drawback, limiting the amount of soluble sugars that can be produced unless they are frequently removed. To degrade cellobiose into glucose first the non-cellulosomal β-glucosidase BglB from *C. thermocellum* was fused with a dockerin module from endoglucanase CelA. The backbone for dockerin-fusion vector (Fig. 34) was the pQE-32 vector. The dockerin module of endoglucanase Cel8A was cloned at the 3`-end of the multiple cloning site.

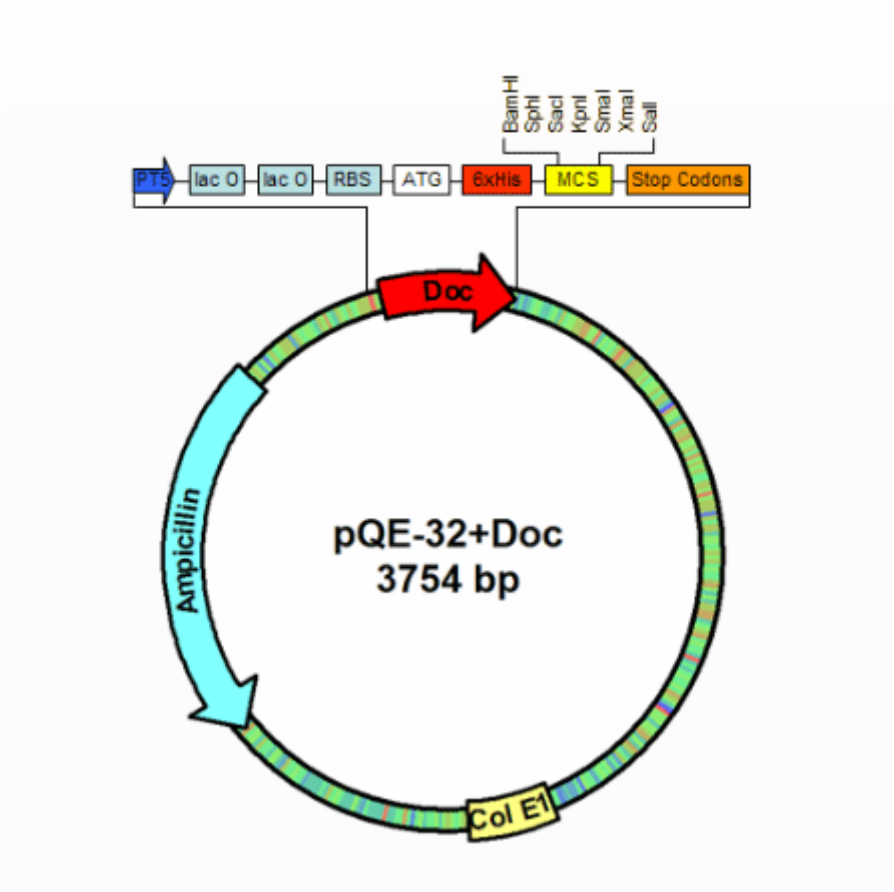


Fig. 34: Dockerin-fusion vector variant pQE-32+Doc.

However, a stability assay of dockerin-fused recombinant BglB showed that relative activity was diminished to ~25% after 40 minutes incubation time at 60°C (Fig. 35). The result is in accordance with the results from Romaniec (1992). The cellobiase activity of the recombinant produced BglB had a half-life of 10 h at 45°C (Romaniec et al., 1992).

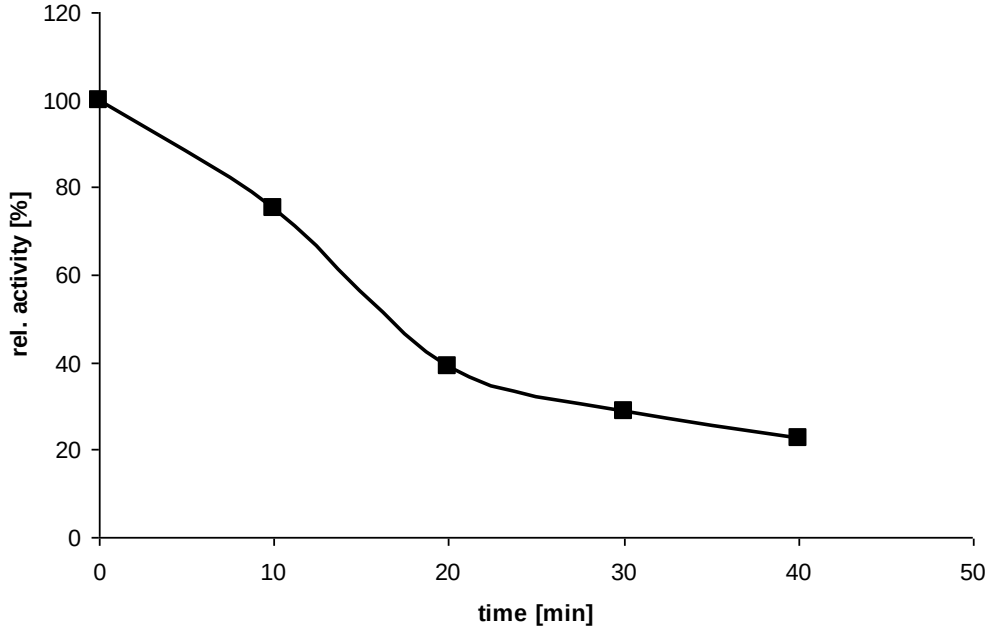


Fig. 35: Relative activity [%] of recombinant *C. thermocellum* BglB fused with dockerin. 10 µg of recombinant BglB fused with dockerin was incubated with 4-nitrophenyl(PNP)-β-glucoside for 40 minutes at 60°C.

Alternatively a thermostable β-glucosidase BglB from *Thermotoga neapolitana* was modified with dockerin-fusion vector by C-terminal fusing with a dockerin module of endoglucanase Cel8A from *C. thermocellum*. Thus, the enzyme can interact specific with cohesin bearing miniscaffoldins. The purified recombinant β-glucosidase (BglB-Doc) had a molecular mass of 92.6 kDa (Fig. 36, A) in accordance with the amino acid sequence. The enzyme had a specific activity of 255 U/mg on 4-nitrophenyl(PNP)-β-glucoside at the optima of pH 6.0 and temperature of 90°C. To guarantee functionality to cohesin-interaction of fused dockerin non-denaturing PAGE with complexed components (BglB-Doc and two cohesins containing miniscaffoldin) was carried out (Fig. 36, B). BglB-Doc was mixed with CBM-cohesin3-cohesin4, and complex formation took place with miniscaffoldin. The stoichiometric mixtures of the enzyme and miniscaffoldin resulted in a single band with altered mobility, thus indicating that complete or near complete complexation was achieved.

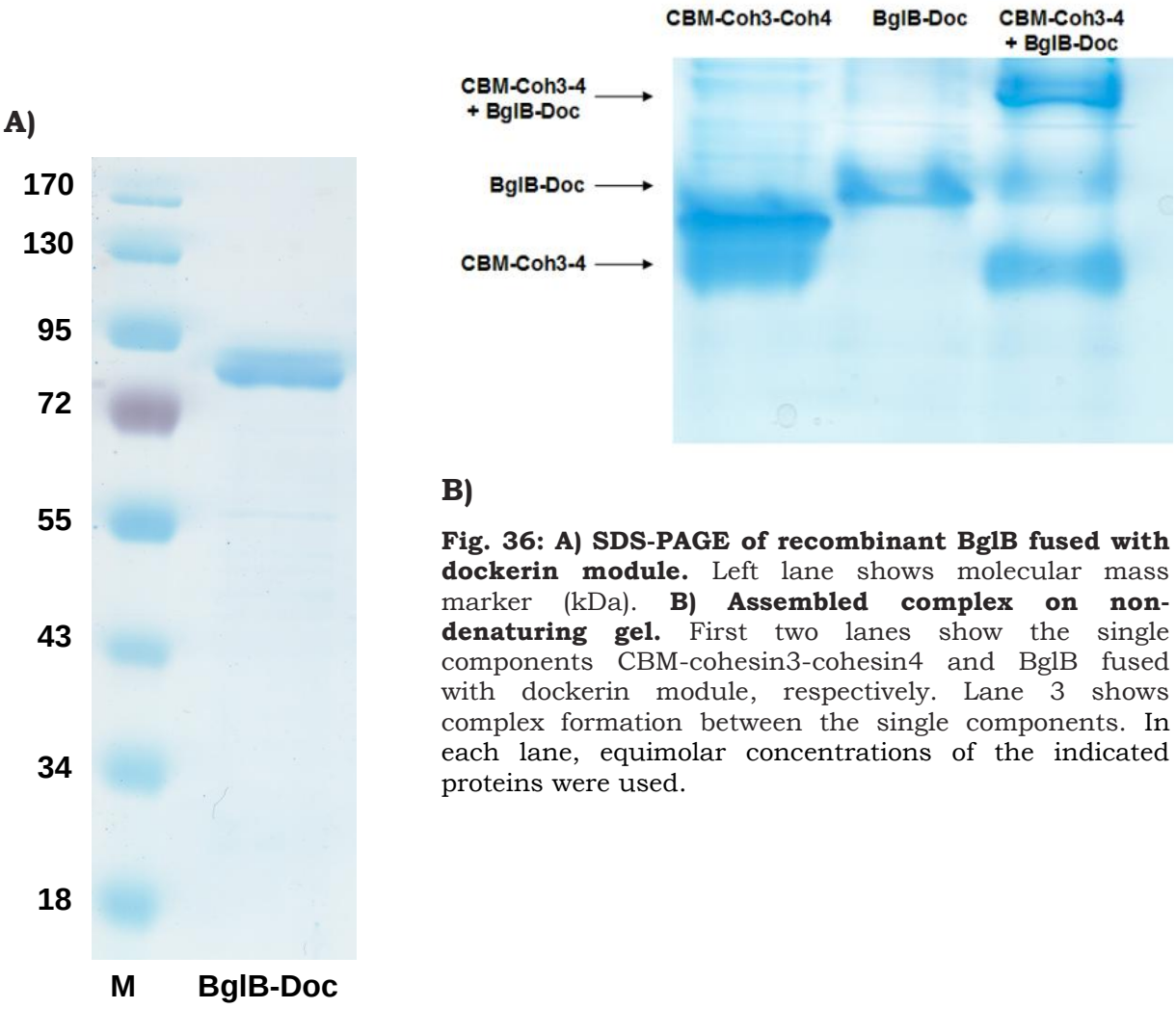


Fig. 36: A) SDS-PAGE of recombinant BglB fused with dockerin module. Left lane shows molecular mass marker (kDa). **B) Assembled complex on non-denaturing gel.** First two lanes show the single components CBM-cohesin3-cohesin4 and BglB fused with dockerin module, respectively. Lane 3 shows complex formation between the single components. In each lane, equimolar concentrations of the indicated proteins were used.

The effect of β-glucosidase BglB is apparent if the release of glucose by degradation of 0.5% Avicel by SM901 mutant enzymes is compared with the activity of SM901 mutant proteins mixed together with BglB (Fig. 37). SM901 mutant enzymes alone as well as those mixed with BglB showed a typical hyperbolic curve progression. After ~ 80 hours (S_1), the release of glucose by degradation of Avicel by SM901 mutant enzymes ceased, allowing for the assumption that product inhibition occurred. In contrast, the curve did not saturate until ~ 140 hours (S_2) if β-glucosidase was added to the enzyme mixture. Thus product inhibition was diminished by the degradation of cellobiose to glucose. Additionally, the amount of released glucose of samples with added BglB is twice as without the enzyme. This indicates that cellobiose was degraded by the enzyme into glucose, resulting in nearly twice the number of reducing ends.

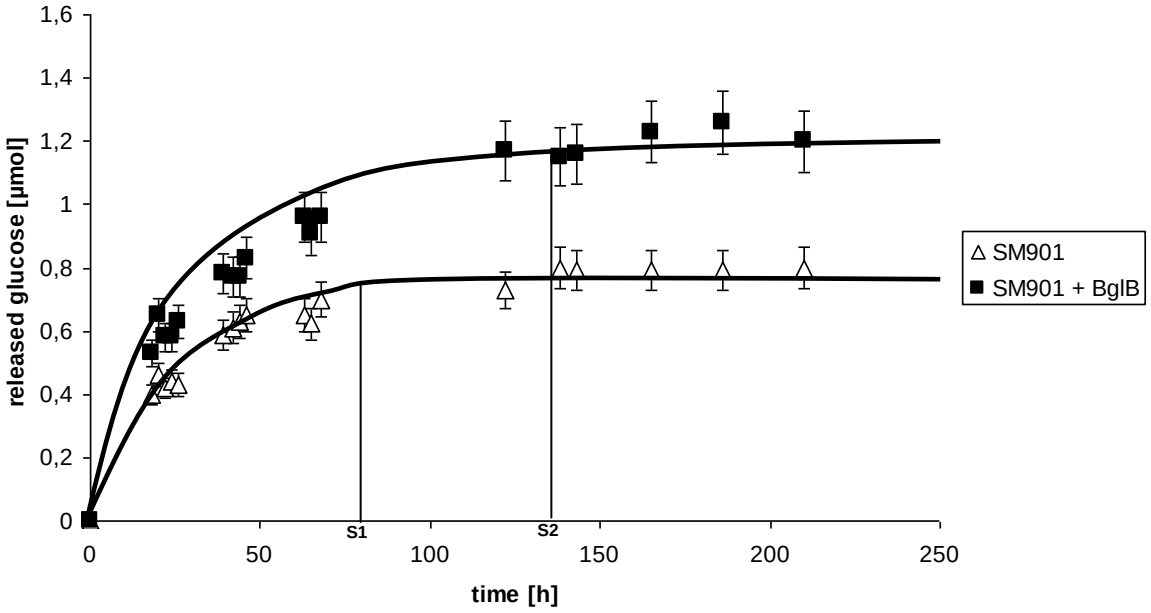


Fig. 37: Release of glucose [µmol] by degradation of Avicel. 10 µg of SM901 mutant hydrolytic enzymes and 10 pmol of recombinant BglB were applied. Samples were incubated at 60°C for 220 hours. At different time points measurements were taken. Amount of released glucose was determined by PGO-test. Curves are labelled as follows: SM901 mutant enzymes are represented by open triangles (Δ), SM901 proteins and additional BglB by filled squares (■), (S_1, S_2) points of curve saturation.

4.4.4 Xylanase Xyn11A, Xyn10C, Xyn10Z and Xyloglucanase Xgh74A

The major carbohydrates making up the primary plant cell wall are cellulose, hemicellulose and pectin. The cellulose microfibrils are linked via hemicellulosic tethers to form the cellulose-hemicellulose network, which is embedded in the pectin matrix.

Hemicellulose is a polysaccharide related to cellulose that comprises approximately 20 % of the biomass of most plants. In contrast to cellulose, hemicellulose is derived from several sugars in addition to glucose, including primarily xylose but also mannose, galactose, rhamnose, and arabinose. Hemicellulose consists of shorter chains, around 200 sugar units. Furthermore, hemicellulose is branched, unlike cellulose.

For effective dismantling of natural substrates it is necessary to introduce xylanases into the artificial complexes. Three xylanases and one xyloglucanase were produced recombinantly and investigated for synergistic action by complex formation with miniscaffoldins.

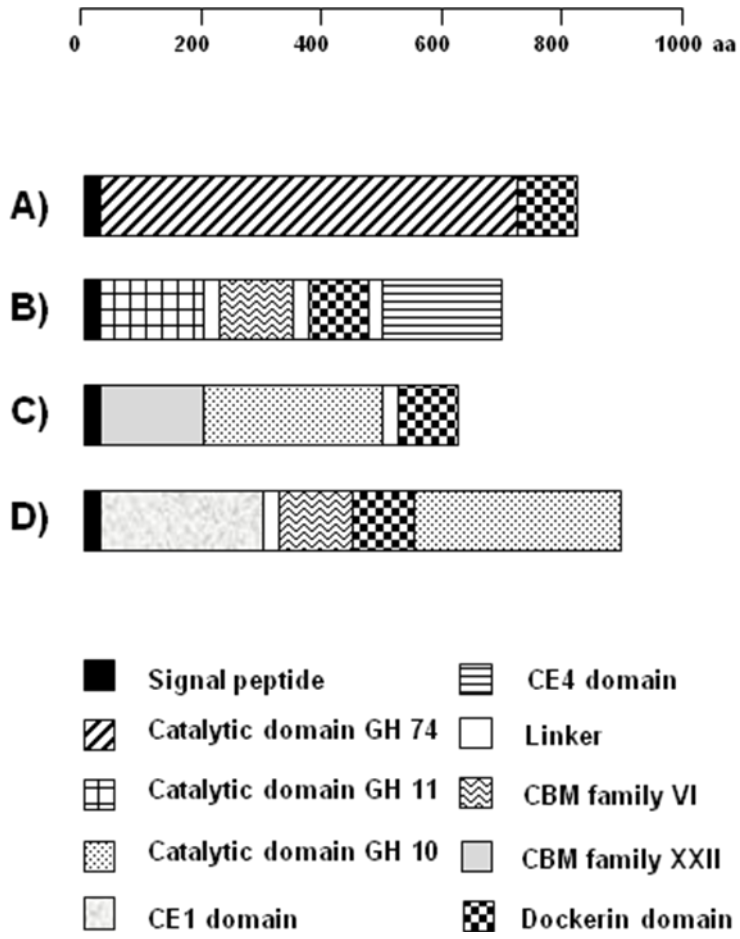


Fig. 38: Domain structure of native xylanases and xyloglucanase. A) Xyloglucanase Xgh74A **B)** Xylanase Xyn11A **C)** Xylanase Xyn10C **D)** Xylanase Xyn10Z

The modular structures of the xylanases and the xyloglucanase are shown in figure 38. The recombinantly produced xylanases and the xyloglucanase are able to interact with cellulosomal CipA protein. All of them consists among other things of a dockerin type I module for interaction with cohesins of CipA. Xylanase Xyn11A and Xylanase Xyn10Z display their family specific catalytic module and a carboesterase module, CE4 and CE1, respectively. Additionally all xylanases exhibit a carbohydrate binding module (CBM family VI and XXII) for interaction with substrate. The CBM VI module is able to bind unsoluble xylan. However the hydrolytic activity on soluble xylan is not affected by present CBM VI (Fernandez et al., 1999). The CBM XXII could act as thermo-stabilising domain. Previous studies showed that removal of this modlue decreases thermo stabilisation drastically (Hayashi et al., 1997).

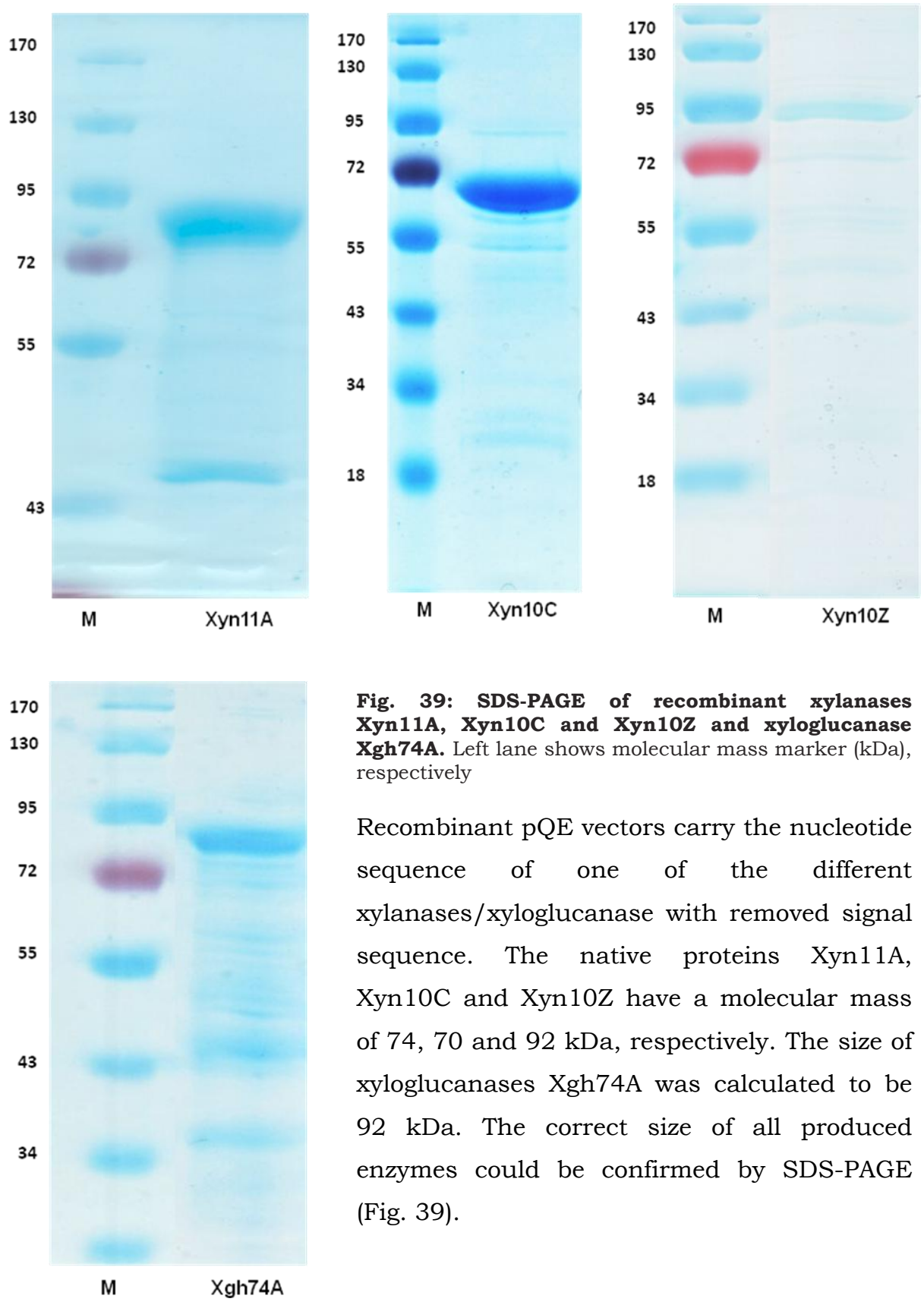


Fig. 39: SDS-PAGE of recombinant xylanases Xyn11A, Xyn10C and Xyn10Z and xyloglucanase Xgh74A. Left lane shows molecular mass marker (kDa), respectively

Recombinant pQE vectors carry the nucleotide sequence of one of the different xylanases/xyloglucanase with removed signal sequence. The native proteins Xyn11A, Xyn10C and Xyn10Z have a molecular mass of 74, 70 and 92 kDa, respectively. The size of xyloglucanases Xgh74A was calculated to be 92 kDa. The correct size of all produced enzymes could be confirmed by SDS-PAGE (Fig. 39).

4.5 Immobilization on surface modified nanoparticles

In recent years, nanoscale materials demonstrated a great potential to serve as superior enzyme supports due to their large surface-to-volume ratio in comparison to traditional macroscale materials. In particular, use of magnetic nanoparticles as a support for immobilized enzymes has achieved growing attention because of the following advantages: a higher specific surface area obtained for the binding of a larger amount of enzymes; lower mass transfer resistance; and selective separation of immobilized enzymes from a reaction mixture by application of a magnetic field.

In this study miniscaffoldins with different quantities of cohesins and optionally a carbohydrate binding module (CBM) were immobilized on surface modified nanoparticles. Several experiments showed that direct immobilization of hydrolytic enzymes on nanoparticles diminished their specific activity. This indicates that active or structural domains of enzymes could be affected by direct covalent coupling. A directional specific coupling of miniscaffoldins was therefore chosen. This reaction guarantees binding of dockerin bearing enzymes to coupled protein carriers via cohesin interaction without loss of activity caused by immobilization.

Nanoparticles (Tab. 10) were purchased from Estapor Microsphere (Fontenay, France). They have a mean diameter of ~100 nm and exhibit superparamagnetic behaviour. The surface of particles is modified either with free carboxyl groups or with free amino groups for coupling reaction with crosslinking agents.

Modification	Mean diameter [µm]	Solid content [%]	Surface groups [µeq/g]
NH_2	0.106 ± 0.007	5	11
COOH	0.110 ± 0.007	9	497

Tab. 10: Characteristics of used nanoparticles.

Different miniscaffoldins were immobilized with interconnected heterobifunctional PEG-based linkers on either NH_2- or COOH-modified nanoparticles (NP). Some differences concerning coupling efficiencies and dispersion abilities occurred.

Modification	Glutaraldehyde [mM]	EDC/Sulfo-NHS [mM]	Coupling efficiency [µg/mg NP]
NH_2	0.5		38.2 ± 3.6
	1		49.7 ± 3.1
	2.5		61.7 ± 3.8
	5		58.3 ± 4.2
COOH		1/2	57.7 ± 3.0
		2/2	63.8 ± 4.4
		2/5	79.6 ± 4.5
		5/5	64.7 ± 4.1
		5/10	60.2 ± 4.5

Tab. 11: Coupling efficiency at different amount of crosslinking agents.

Coupling efficiency (Tab. 110) was determined by spectrophotometric measurement of the optical adsorption (590 nm, Bradford assay) of protein content before and after crosslinking. Protein loaded nanoparticles were separated from reaction solution using a strong disc magnet (Fig. 33). The coupling efficiency was calculated by subtraction the remaining protein in reaction solution from the primary applied amount of protein. Best results were obtained with 2.5 mmol glutaraldehyde for NH_2-modified nanospheres and 2 mmol EDC, 5 mmol Sulfo-NHS for COOH-modified nanospheres, respectively. The maximal coupling efficiency of 80 µg/mg NP could be achieved with COOH-modified particles. This corresponds to a calculated average number of ~ 1300 protein carrier per particle. However, the efficiency with NH_2-modified particles averaged just 60 µg/mg NP. Due to the usage of the homobifunctional crosslinker glutaraldehyde, amino modified beads reacted with free amino groups of other present particles and some microscopic visible aggregation occurred. No crosslinking between carboxy

modified particles could be observed. Additionally, caused by the free carboxy groups and the resulting hydrophilic surface, COOH-beads exhibit much easier resuspendation abilities.

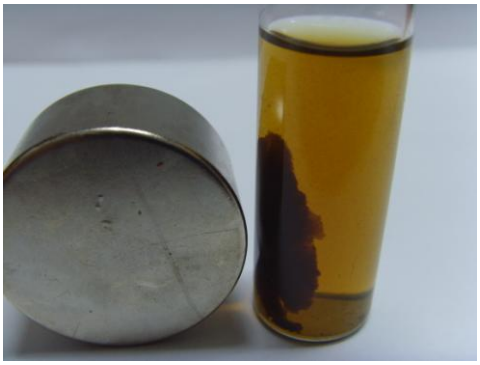


Fig. 40: Magnetic nanoparticles are orientated towards the applied magnetic field.

Due to the superparamagnetic behaviour the particles could be separated from solution by application of magnetic field using a strong disc magnet (Fig. 40). The particles orientated towards the magnetic field and reaction solution was removed. The recovery rate of separated particles was between 93 to 97 %. For regeneration of the magnetite nanoparticles particles were washed three times with EDTA (10 mM) to remove complexed Ni^{2+}-ions and suspended in the same solution overnight at room temperature. After three washing steps with deionised water the particles could be reloaded with Ni^{2+} and 6xHis-tagged miniscaffoldins. 1 mg of fresh COOH modified nanoparticles could be loaded with ~ 80 µg protein carrier. Recycled nanoparticles of the same amount were able to bind ~ 50 µg (62.5 % recovery) of miniscaffoldins, and twice recycled nanoparticles could immobilize ~ 25 µg (31.2 % recovery) protein (Tab. 12). NH_2-modified particles showed a diminished initial charge density as well as recovery rate in contrast to COOH-modified beads.

Modification	Initial charge density	First recharge	Second recharge
		[µg protein/mg NP]	
NH_2	61.7 ± 3.8	37.5 ± 3.9	17.2 ± 2.8
COOH	79.6 ± 4.5	51.3 ± 3.7	24.6 ± 3.2

Tab. 12: Recharge efficiency [µg/mg NP] of used nanoparticles.

4.6 Specific activities of constructed complexes on different substrates

Activity assays were performed with immobilized miniscaffoldins differing in the amount of cohesins as well as the presence of carbohydrate binding module. While the simplest miniscaffoldin is composed of only a single cohesin, the most complex consists of 3 different cohesins and a carbohydrate binding module. Hydrolases were immobilized by cohesin-dockerin recognition on the miniscaffoldin-nanoparticle complex. To investigate synergistic behaviour between neighbouring hydrolases, complexes containing two cohesins were compared with complexes bearing just a single cohesin. The effect of CBM in complex was further investigated by contrast miniscaffoldins of two cohesins bearing a CBM or not.

Initially the constructed complexes were loaded with SM901 mutant enzymes and their specific activity towards soluble, amorphous and insoluble cellulose was determined.

Further, two different amount and combination of recombinant cellulases were added to SM901 mutant enzymes. The relevant hydrolases were first mixed with SM901 mutant enzymes stoichiometrically according to previous studies from Gold et al. (2007) to balance the disadvantage of on soluble cellulose obtained mutant enzymes reffering to the composition of the hydrolases in a potential cellulosome (which was not produced by the mutant) to hydrolysis of crystalline cellulose. Secondly, the recombinant enzymes were immobilized without any mutant enzymes to identify the components, which are necessary for effective degradation of crystalline cellulose.

Thirdly, the recombinant β-glucosidase BglB from *Thermotoga neapolitana* fused with a dockerin module that was added to enzyme mixtures to prevent product inhibition due to the main product cellobiose. The addition of complementable components from another cellulase system and organism was possible due to the usage of the pQE-dockerin-fusion vetcor.

References for activity tests were always the free mutant enzymes and the native cellulosome of *C. thermocellum*.

4.6.1 Hydrolytic activity on soluble and amorphous cellulose

As previously reported, the SM901 mutant enzymes showed no significant differences in degradation rate between soluble substrates and the native cellulosome of *C. thermocellum* (Zverlov at al., 2008). Also the complexed SM901 hydrolases bearing miniscaffoldins exhibited a similar specific activity towards soluble cellulose with no significant differences. This demonstrated that the complexation and direct proximity of hydrolytic enzymes to each other do not play a major role in the degradation of non-crystalline substrates. The following figures show the specific activities towards β-glucan, carboxymethyl cellulose (CMC) and phosphoric acid swollen cellulose.

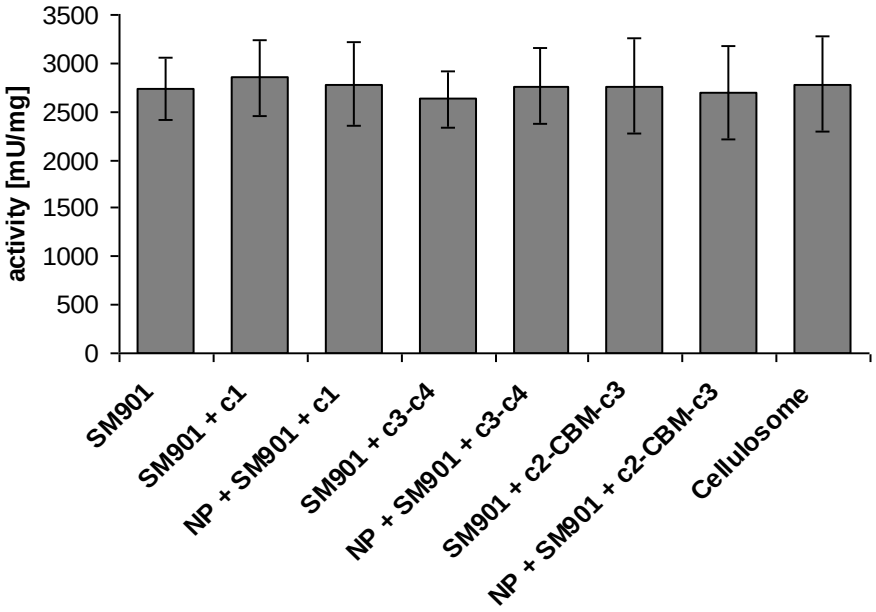
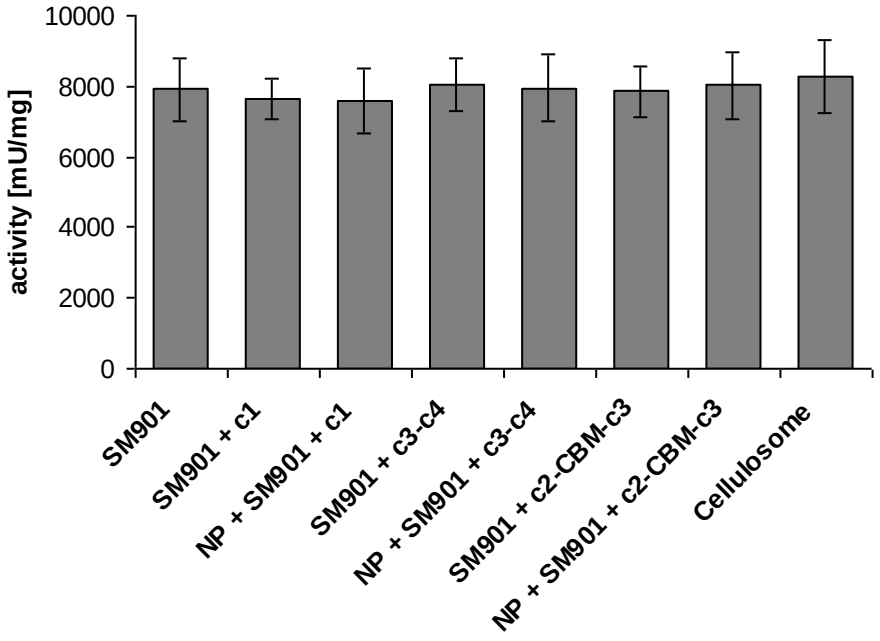


Fig. 41: Specific activity [mU/mg protein] of mutant enzymes SM901, in complexed form on nanoparticles with different miniscaffoldins and native purified cellulosome on 0.5 % PASC. CBM, carbohydrate binding module; c, cohesin; NP, nanoparticles.

However, the effectiveness of degradation is dependent on the type of substrate. Most accessible for hydrolysis is barley β-glucan. The specific activity towards this substrate was determined of approximately 8 U/mg protein (Fig. 42 A). The degradation of carboxymethyl cellulose by the exoglucanases of SM901 mutant enzymes is impeded by exposed carboxymethyl groups resulting in a hydrolytic rate of about 1.1 U/mg protein (Fig. 42 B). Amorphous cellulose is distinguished by exposed cellulose fibers among crystal structure. These amorphous regions are more

accessible for degradation, resulting in specific activities of about 2.8 U/mg protein (Fig 41). The immobilization of hydrolytic enzymes on nanoparticles connected miniscaffoldins had no negative effect on the degradation rate for all tested substrates.

A)



B)

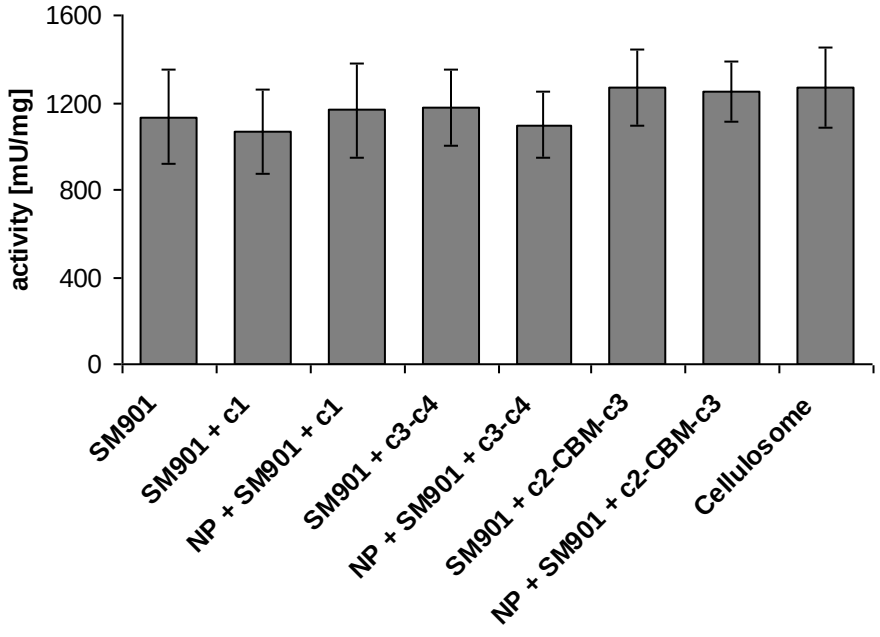
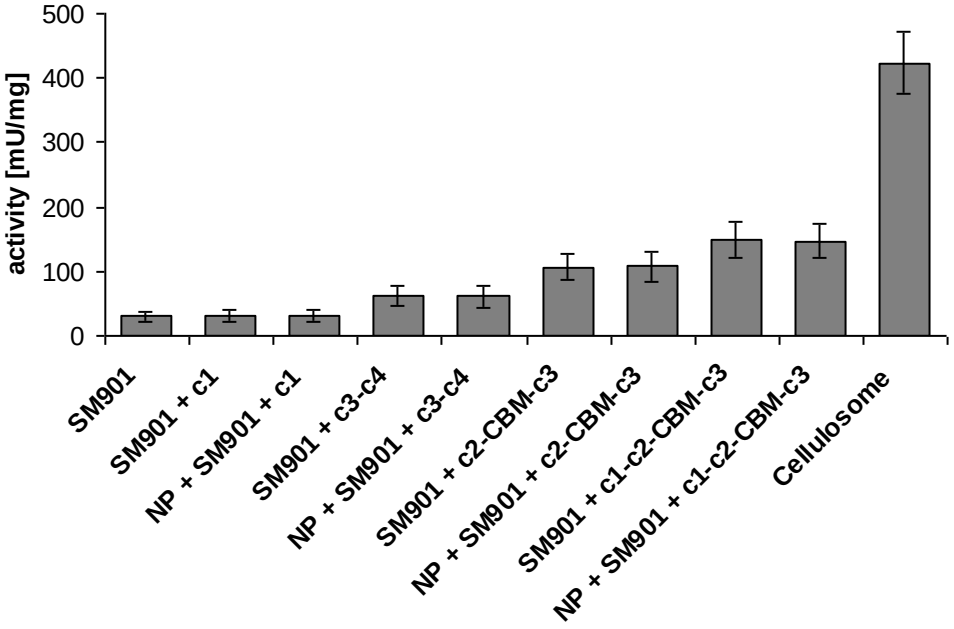


Fig. 42: Specific activity [mU/mg protein] of mutant enzymes SM901, in complexed form on nanoparticles with different miniscaffoldins and native purified cellulosome on soluble 0.5 % barley β-glucan (A) and CMC (B). CBM, carbohydrate binding module; c, cohesin; NP, nanoparticles.

4.6.2 Hydrolytic activity on insoluble cellulose

On crystalline cellulose, specific activities are clearly lower than those of soluble cellulose (Fig. 43). The substrate exhibits a crystalline structure and is therefore much less susceptible to enzymatic degradation. Free mutant enzymes showed a specific activity on MN300 and Avicel of 30 and 12 mU/mg respectively. Purified native cellulosome with 9 cohesins and a carbohydrate binding module exhibited a specific activity of 423 mU/mg with MN300 as substrate, and 198 mU/mg with Avicel. The complexation of SM901 mutant enzymes with a single cohesin seems to bring neither advantage nor disadvantage in the degradation of crystalline cellulose. Specific activity remained constant on both substrates. With increasing quantities of cohesins in miniscaffoldin the hydrolytic activity was enhanced by the close proximity of two or more enzymes. Bringing two enzymes in accidental combination into close contact by complexation with two cohesin bearing miniscaffoldin resulted in specific activities of 63 and 28 mU/mg for the two substrates. The enhancement of the activity was 2.1 and 2.3-fold over free enzymes. If a complex with three cohesins was used, the degradation rate would increase 4.9 and 3.7-fold compared to unbound hydrolases. However, this complex contained a family-3 carbohydrate binding module, which is known to interact strongly with crystalline cellulose. By comparison miniscaffoldins, containing a CBM or without CBM (cohesin3-cohesin4 and cohesin2-CBM-cohesin3) the specific activity increased from 62 mU/mg to 102 mU/mg with MN300 as the substrate and 44 mU/mg to 108 mg/mU with Avicel if a CBM was present in miniscaffoldin. To investigate whether the position of a CBM within miniscaffoldin has an impact on synergistic behaviour, a miniscaffoldin with CBM positioned between two cohesins (cohesin2-CBM-cohesin3) was compared to a miniscaffoldin containing two cohesins and a CBM at the end position (CBM-cohesin3-cohesin4). The specific activities of mutant enzymes with these two miniscaffoldins did not vary, indicating that the order of cohesins and CBM did not play a major role in dismantling crystalline cellulose. The immobilization of hydrolytic enzymes on nanoparticle-bound miniscaffoldins had no negative effect on degradation rate of all tested substrates.

A)



B)

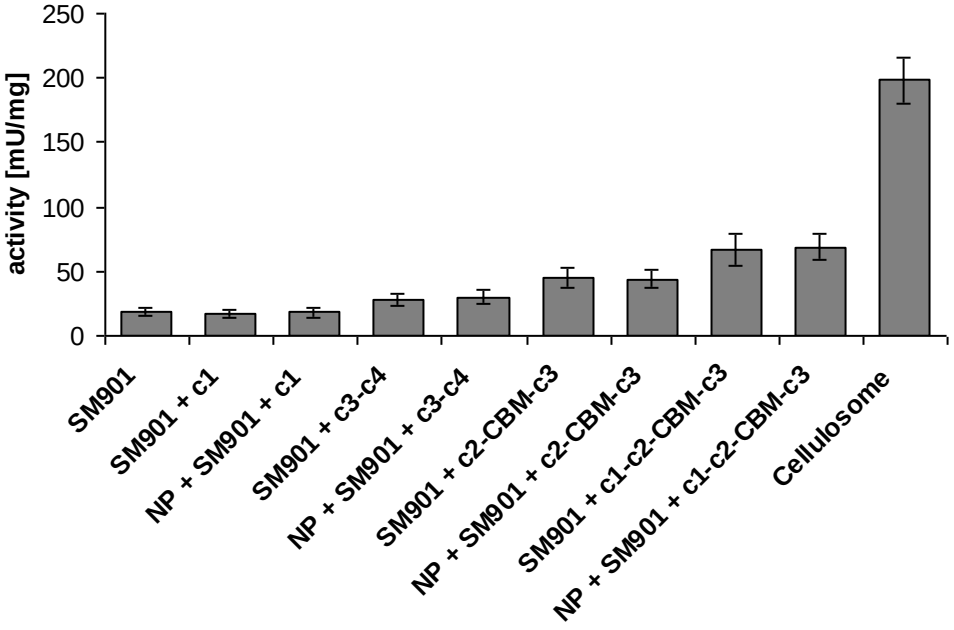


Fig. 43: Specific activity [mU/mg protein] of mutant enzymes SM901, in complexed form on nanoparticles with different miniscaffoldins and native purified cellulosome on 0.5 % MN300 (A) and Avicel (B). CBM, carbohydrate binding module; c, cohesin; NP, nanoparticles.

4.6.3 Hydrolytic activity on insoluble cellulose with additional enzymatic components

The exoglucanases CbhA, CelK and CelS and endoglucanases CelJ and CelR were added to SM901 mutant enzymes in accordance with the studies of Gold et al. (2007), both with and without scaffoldin protein CipA. Native cellulosomes purified from *C. thermocellum* were always the reference. The components were then mixed to mutant proteins in different amounts and combinations. Avicel was chosen as the model substrate, due to its highest crystalline structure. However, CipA could not be immobilized on nanoparticles due to insufficient recovery from *E. coli* host. Further studies showed that immobilization of miniscaffoldins on nanoparticles has no negative effect on hydrolytic activity (see section 4.6.2). It could therefore be hypothesized that scaffoldin CipA proteins not immobilized show equal behaviour toward degradation of crystalline cellulose as if they were immobilized on surface of nanoparticles. Initially, hydrolytic activity on Avicel of the single glucanases in free unbound state and complexed with scaffoldin protein CipA were determined (Fig. 44).

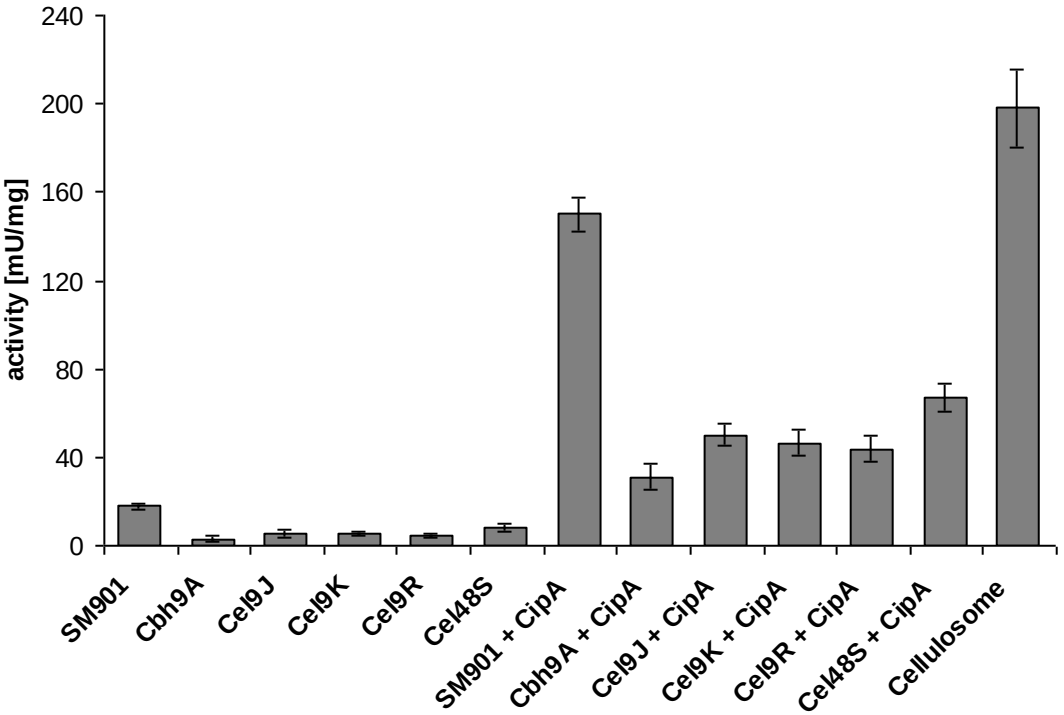


Fig. 44: Specific activity [mU/mg protein] of mutant free enzymes SM901, the single recombinant exo- and endoglucanases and bound to scaffoldin protein CipA. All samples were tested on 0.5 % Avicel and compared with native purified cellulosome.

Recombinantly produced glucanases showed diminished activities between 3.1 and 8.3 mU/mg compared to the SM901 mutant enzymes (18.3 mU/mg). This indicates that a single type of enzyme does not show the strong hydrolytic activity as the same amount of mixture of different enzyme types. Enzyme mixtures with scaffoldin protein CipA were generally found to be more active than simple mixtures of the free enzymes. The degradation rate increased 8- to 9-fold due to synergistic characteristics caused by their closed proximity and the presence of CBM in the scaffoldin protein. Furthermore, the different recombinant hydrolytic enzymes were added to the SM901 mutant cocktail in different amounts and combinations (Fig. 45).

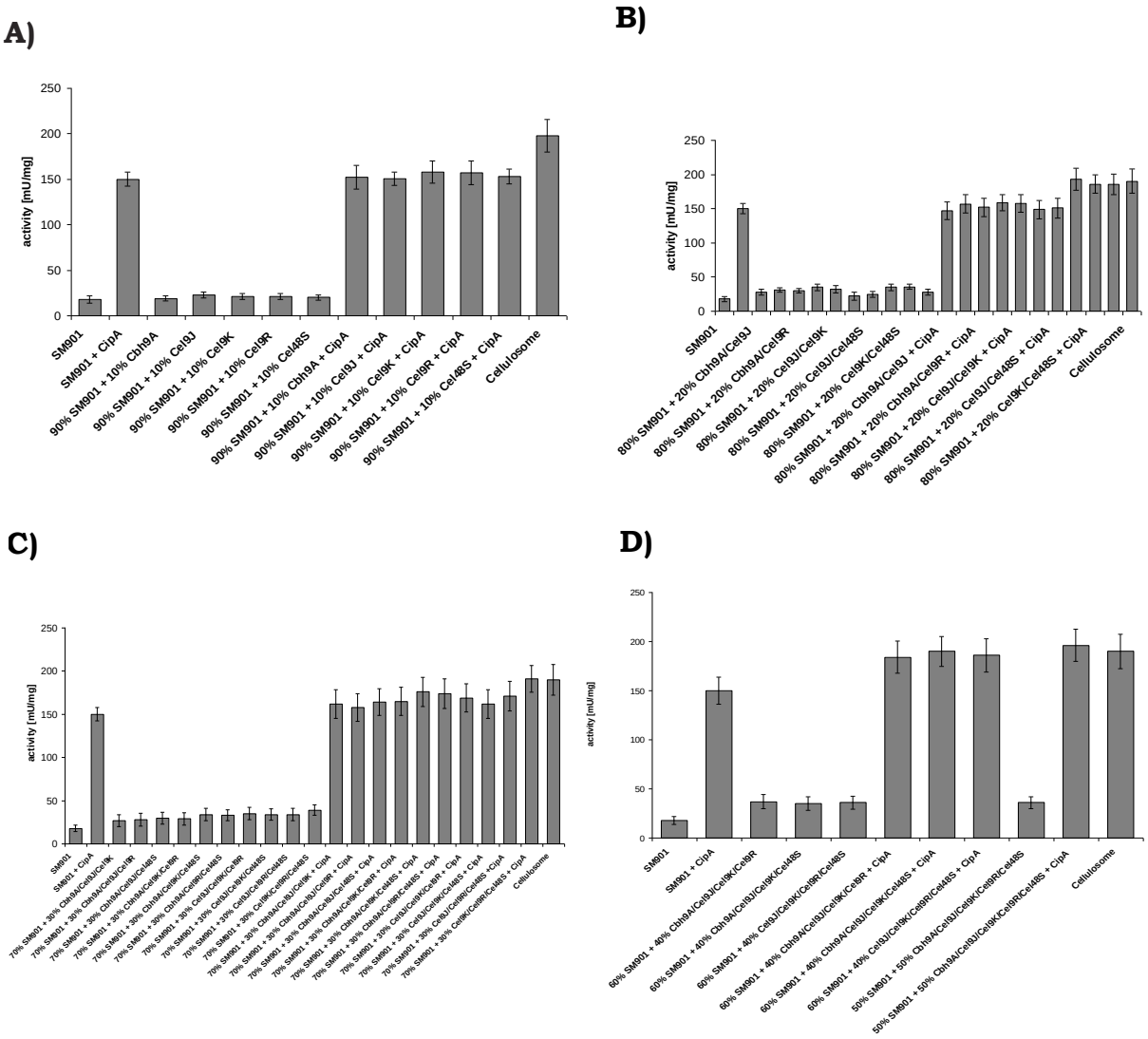


Fig. 45: Specific activities [mU/mg] of mutant enzyme cocktail mixed with different amount and types of recombinant hydrolases. A) 90 % (mol) of SM901 mutant enzymes mixed with 10 % (mol) Cbh9A, Cel9J, Cel9K, Cel9R and Cel48S, respectively. **B)** Two, **C)** three, **D)** four and five different recombinant hydrolases are mixed together in varying combination and molar ratio to SM901 hydrolases. All samples were tested on 0.5 % Avicel and compared with native purified cellulosome.

The results clearly demonstrated that the addition of underrepresented compounds of the on cellobiose grown SM901 mutant to enzyme cocktail increased hydrolytic activity towards high crystalline cellulose. A specific activity of 18.1 mU/mg was determined for mutant enzymes. After stoichiometric addition of 10 % Cel9K, Cel9R, Cel9J and Cel48S each, the production rate of glucose increased to 37 mU/mg, representing the most effective combination. Other combinations produced similar results with activities between 20 and 35 mU/mg.

4.6.4 Hydrolytic activity of complexed xylanases/xyloglucanase

Activity of recombinant xylanases and the xyloglucanase were determined with oat spelt xylan as substrate by DNSA test. The substrate was fractionated in soluble and insoluble components through heating and a series of washing steps. All xylanases showed hydrolytic activity towards soluble xylan, whereas Xyn11A exhibited the highest degradation rate of ~ 685 U/mg followed by Xyn10C with 119 U/mg and Xyn10Z with 58 U/mg. However, the xyloglucanase Xgh74A showed no activity towards this substrate (Fig. 46). Xgh74A did, however, show strong hydrolytic activity on barley β-glucan, in agreement with the studies of Zverlov et al. (2005).

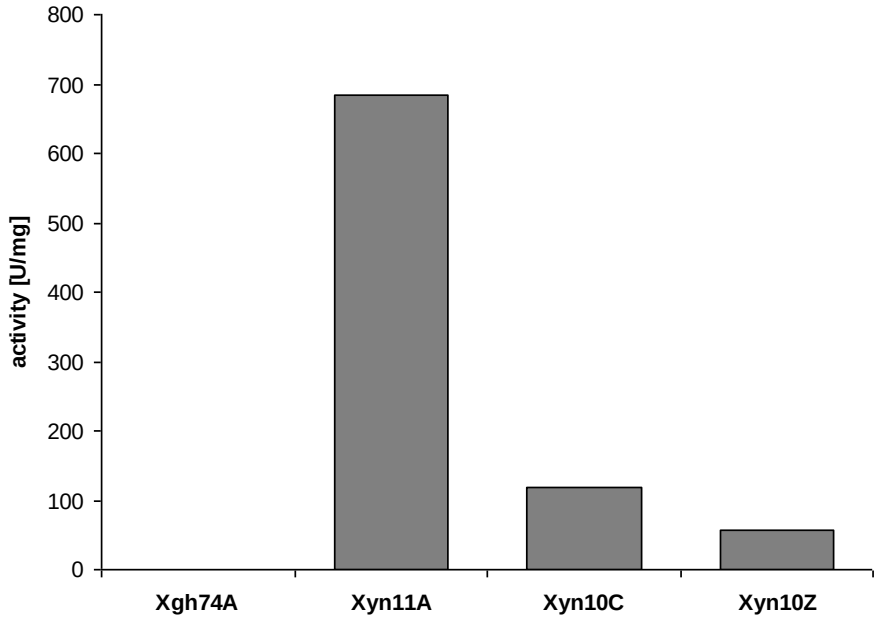


Fig. 46: Specific hydrolytic activities [U/mg] of recombinant xylanases Xyn11A, Xyn10C, Xyn10Z and xyloglucanase Xgh74A on soluble fraction of oat spelt xylan.

The strongest hydrolytic activity on insoluble components of oat spelt xylan could also be achieved with the xylanase Xyn11A with 1.22 U/mg. The specific activities of xylanase Xyn10C and xylanase Xyn10Z were 1.01 and 0.68 U/mg, respectively. After complexation with miniscaffoldin type CBM-Coh3-Coh4 (c-3-4), a distinct increase in degradation could be observed (Fig. 47). On average, the hydrolysis rate could be increased by ~10%.

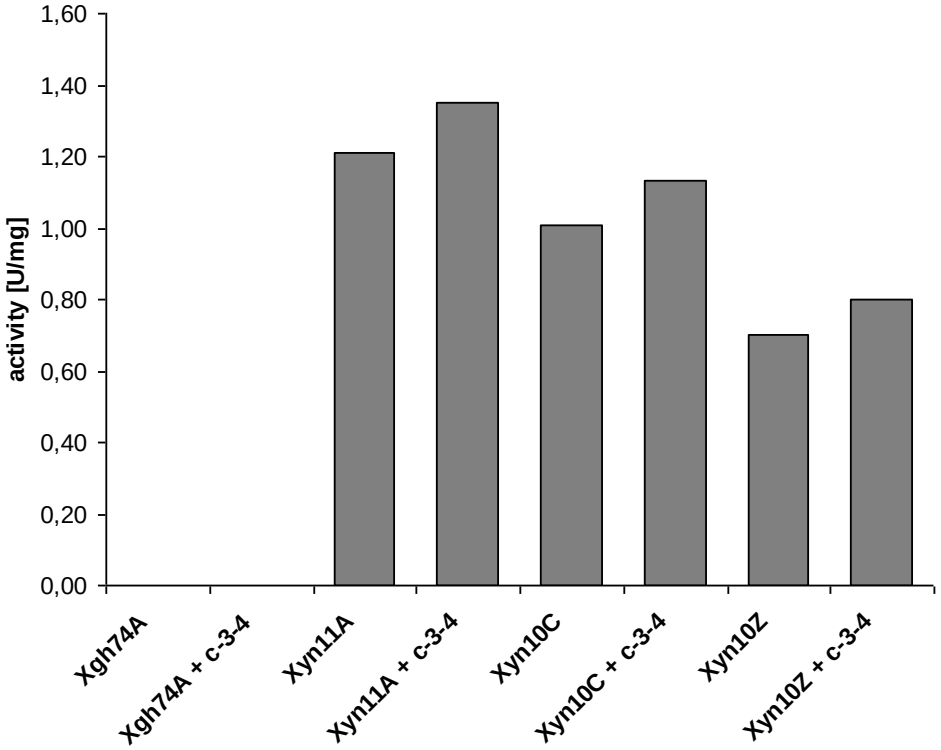


Fig. 47: Specific hydrolytic activities [U/mg] of single recombinant xylanases Xyn11A, Xyn10C, Xyn10Z and xyloglucanase Xgh74A and mixed with miniscaffoldin CBM-cohesin3-cohesin4 (c-3-4) on insoluble fraction of oat spelt xylan.

Distinctly better results could be achieved by mixing two different components together with miniscaffoldin c-3-4 (Fig. 48). The combinations of equal fractions of Xyn11A and Xyn10Z or Xyn10C and Xyn10Z complexed with miniscaffoldin are also noteworthy. An enhancement of hydrolytic activity of 23% and 27% was observed. The xyloglucanase was intentionally ignored because of its known lack of hydrolytic activity on oat spelt xylan.

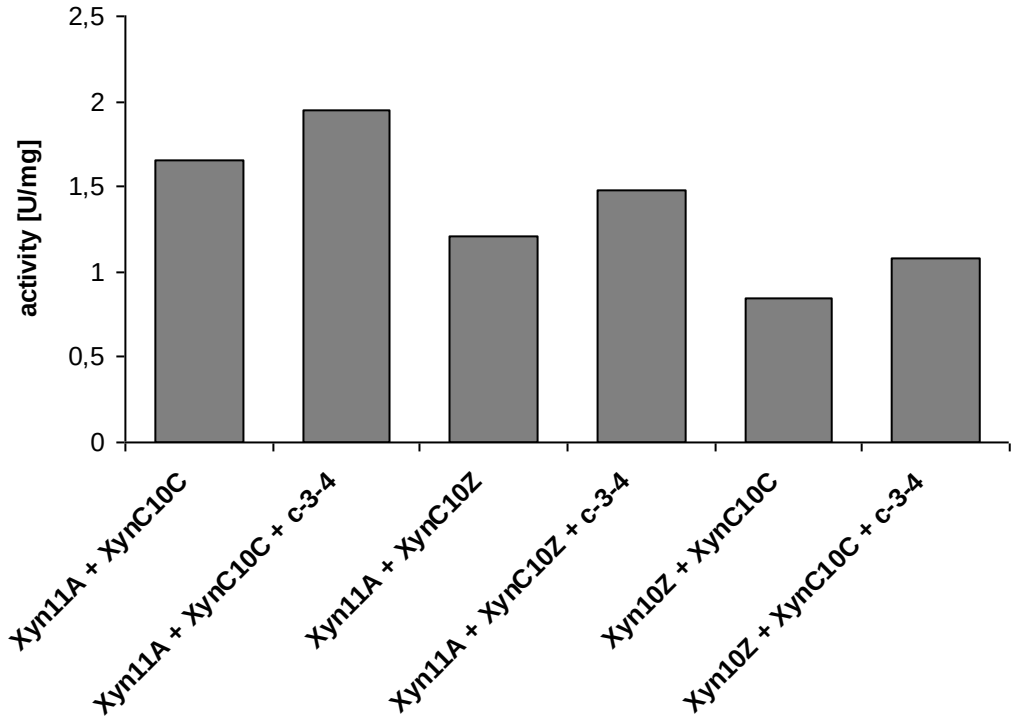


Fig. 48: Specific hydrolytic activities [U/mg] of different mixed recombinant xylanases on insoluble fraction of oat spelt xylan.

4.7 Stability of nanoparticle-miniscaffoldin complexes

It is reported that enzymes immobilized on polystyrene particles (Caruso et al., 2000) or quartz slides (Onda et al., 1999) can enhance their enzyme stability with respect to temperature and pH conditions.

Relative activities of immobilized enzymes on nanoparticles bearing single cohesins and free enzymes were compared at different pH value and temperature (Fig. 49). Free and bound enzymes showed similar temperature distributions (Fig. 49A). The optimum temperature at pH 6.5 corresponding to the highest enzyme activity was observed to be 60 °C. Moving towards higher temperature of 70 °C and 80 °C, more stability could be observed on immobilized enzymes. The relative activity of bound hydrolases was about 68 % at 70 °C, whereas the activity of free once decreased to a value of about 50 % at the same temperature.

At varying pH values between 4 and 9 (T = 60 °C), the immobilized enzymes showed an expanded stability spectrum (Fig. 49B). The optimum pH value was observed to be 7. After shifting into more acidic or alkaline pH ranges, relative activities decreased dramatically. At these extremes the bound

enzymes displayed much more stability in contrast to the free enzymes. At pH ranges from 4-6 the enzyme bearing nanoparticles were about 18 %, 54 % and 150 %, respectively, more active than free unbound SM901 mutant hydrolases.

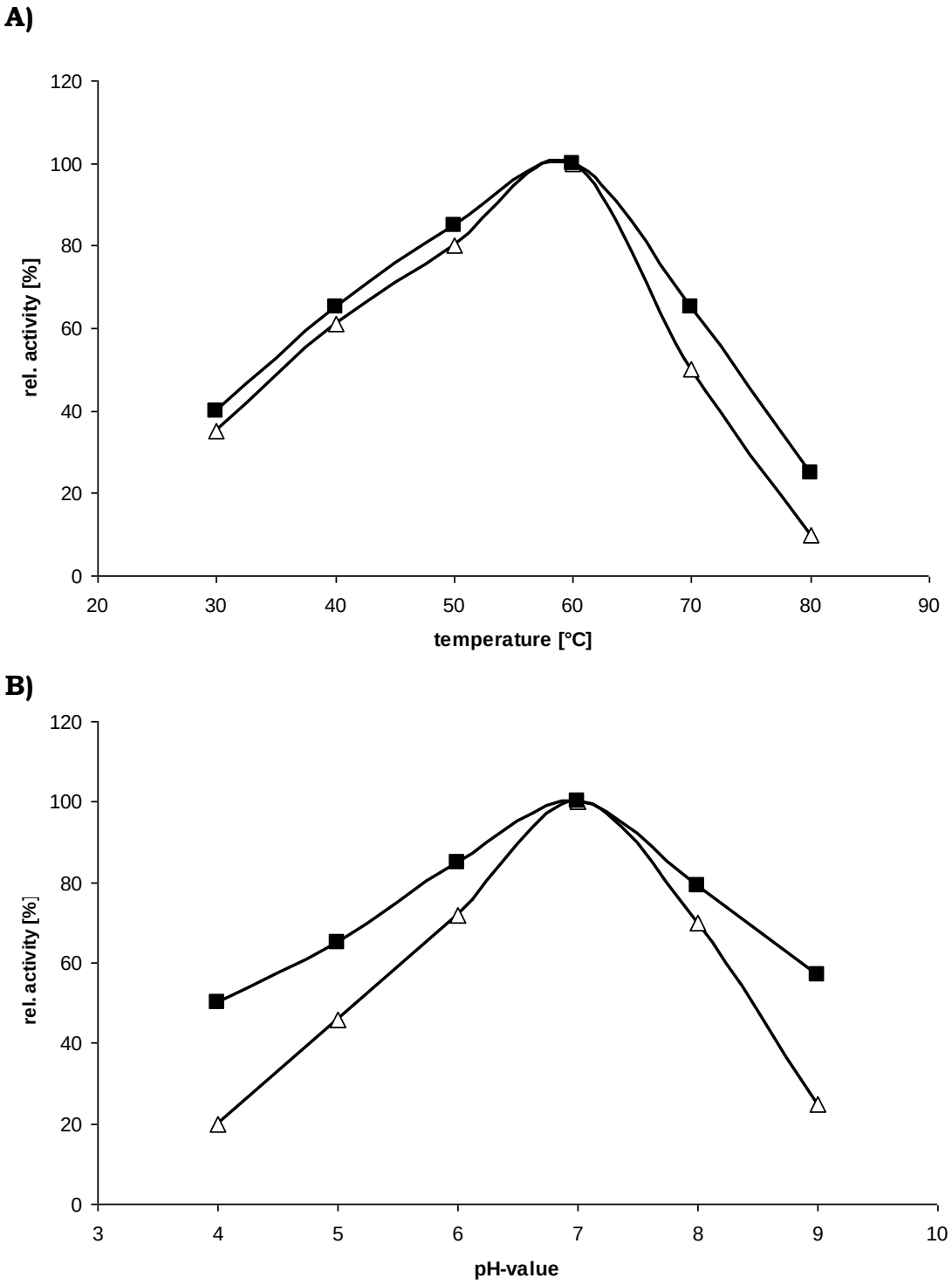


Fig. 49: Relative activities at different temperatures (A) and pH values (B) of free and immobilized SM901 mutant enzymes. Curves are labelled as follows: SM901 mutant enzymes are represented by open triangles (Δ), immobilized on miniscaffoldin c1 bearing nanoparticle by filled squares (■).

Loss of storage stability is a major concern in enzyme preservation. The storage stability of the SM901 exoenzymes was examined for 40 days. Figure 50 shows the storage stabilities of free SM901 mutant enzymes and immobilized enzymes on nanobeads at 60 °C, pH 6.5 (in 0.1 M MES, 0.5 mM $CaCl_2$, 0.1 M NaCl, 0.02 % NaN_3).

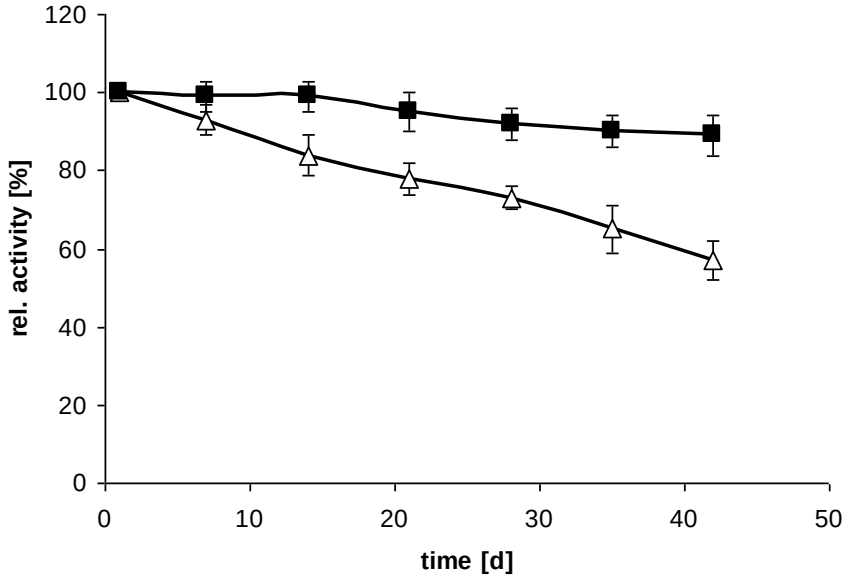


Fig. 50: Relative activity at different time points [d] of free and immobilized SM901 mutant enzymes. Curves are labelled as follows: SM901 mutant enzymes are represented by open triangles (Δ), immobilized on miniscaffoldin c1 bearing nanoparticle by filled squares (■).

Each data point was the average of triplicate measurements. The activity decreased with time in all of the systems. A loss of activity to 60 % was observed after 40 days for the free enzymes, with almost unaltered rates for bound enzymes while enzyme bearing particles retained activity of about 95 % after 40 days of storage under the same conditions. The stability of the enzyme was found to improve upon binding to the magnetic nanoparticles. The fixation on the surface of the magnetic nanoparticles has been a tangible argument supporting the prevention of auto-digestion of the enzyme and lysis, and the subsequent conservation of its activity (Liao et al., 2001). This argument supports these results and justifies the long term stability of enzyme loaded nanoparticles over the free enzymes.

Chapter 5

Discussion

One of the great difficulties in reconstitution of the cellulosome of *Clostridium thermocellum* is the separation of this highly cohesive complex into its component parts. In a previous contribution, Lamed et al. (1983) noted that the cellulosome is surprisingly resistant to proteolysis. Morag et al. (1991) separated the cellulosome by SDS-treatment at elevated temperatures, but such treatment was deleterious to the activity of most of its enzyme components. This prevented successful reconstitution of the cellulosome from native cellulosome components. The detection of the loss of the scaffoldin protein (CipA) in the SM901 mutant culture supernatant allowed a recovery of the complete set of cellulosomal hydrolases in soluble (non-complexed) form and thus an artificial reconstitution of the *Clostridium thermocellum* cellulosome was possible. Moreover the synergistic effect of complexation of the enzymes in a cellulosome could be quantified for the first time. The difference between the complexed and the non-complexed enzymes was 15-fold – an unexpectedly high factor, which gave the project a strong impulse first: to reconstitute the cellulosomal complex artificially, to repeat the synergistic effect *in vitro*, and to reduce the number of components for practical application.
The following acquired aspects and knowledges lead to a successful construction of an artificial cellulosome

- The SDS-PAGE pattern of two different cohesins (1 and 6) loaded with SM901 exoproteins showed that the interaction of dockerin and the corresponding cohesin is not specific within the single cohesins in the scaffoldin CipA protein
- Enzymes bound on miniscaffoldins consisting of two cohesins exhibited a higher specific activity than enzymes in free state; with an

increasing number of cohsesins within a miniscaffoldin a further improvement of hydrolytic performance could be achieved

- A close proximity of the hydrolases to each other leads to synergism
- With the introduction of a carbohydrate binding module (CBM) in the miniscaffoldins, the enzymes are in close contact with the substrate, resulting in an increasement of hydrolytic activity
- With the addition of underrepresented components to the SM901 mutant exoenzymes, the complexes were adapted to an effective hydrolysis of recalcitrant substrates (crystalline cellulose)
- A developed dockerin fusion vector allowed an introduction of user-defined components (naturally without a dockerin module) in the complex
- With the introduction of a thermostable β-glucosidase BglB from *Thermotoga neapolitana* into the miniscaffoldin-nanoparticle complex, inhibition effects by the main product cellobiose could be diminished
- Miniscaffoldins were immobilized on surface modified nanoparticles; their ability to bind dockerin bearing hydrolases was not affected by immobilization
- With the interconnection of a heterobifunctional PEG-based linker between particle surface and miniscaffoldins, the system gets more flexibility
- Due to the magnetic character of the particles, the hydrolytic complexes could be recycled from reaction solution
- Due to the special coupling chemistry disused enzyme bearing miniscaffoldin could be removed from particles and the particles could be recharged with fresh components
- The nanoparticles-enzyme complex exhibited a higher stability to shifted pH and temperature environments compared to non-bound miniscaffoldin-enzyme complexes
- The produced hydrolytic complex exhibited a higher hydrolytic activity than a native cellulosome preparation
- The developed system is a serious competitor for commercially available cellulose preparations

In the following sections the mentioned points are further discussed and evaluated.

5.1 CipA defective *Clostridium thermocellum* mutants

The putative mutated *cipA* genes revealed the insertion of a new identified IS element called IS*1447*. It was inserted in identical orientation in all six mutants investigated, albeit into different positions within *cipA* (Zverlov et al., 2008). This resulted in an interruption of the CipA reading frame and thus in truncated scaffoldins, which were secreted but had limited or abolished complexing ability. The size of the complexes formed in mutants SM906 to SM904 is mirrored in their increasing ability to degrade crystalline cellulose. The reduction of complex size or the entire loss of complexation indicated an inability of the mutants to form integer cellulosomes. The proteins thus appeared in mutant SM901 as dispersed and non-complexed. The SDS-PAGE pattern (Fig. 17) of the proteins in the culture supernatant showed that these proteins resembled the cellulosome components. Only the CipA band was missing entirely. Cellulosome complexes were absent in culture supernatants of mutant SM901. Nevertheless, small complexes with increasing size were detected in the mutants SM904, SM905, and SM906, and some truncated cohesin modules could be produced (Fig. 20) (Zverlov et al., 2008). The mutants of *C. thermocellum* were not impaired in their ability to hydrolyze soluble β-glucan or CMC. The SM901 mutant enzymes did, however, show a ~15-fold decrease in specific activity towards micro-crystalline cellulose in contrast to purified cellulosomes from the wild type (Tab. 3).

The synergism between selected cellulosomal components has been detected earlier to a certain extent. Studies from Bayer et al. (2001) have demonstrated that cellulolytic enzymes incorporated on so-called miniscaffoldins, e.g. truncated parts of CipA, exhibited enhanced synergistic action on crystalline cellulose compared to free hydrolases. The results with the mutant SM901 enzymes are therefore in agreement with the cellulosome paradigm (Fierobe et al., 1998). The cellulosome paradigm attributes the high efficiency of the complex on crystalline cellulose to the high local concentration of synergistic components. This is ensured by the close

proximity of single hydrolases incorporated on the scaffoldin protein, e.g. CipA of *Clostridium thermocellum*.

Synergistic effects of incorporated hydrolases could also be observed in other cellulosome producing Clostridia besides *Clostridium thermocellum*, such as *Clostridium cellulovorans*, *Clostridium cellulolyticum* and *Clostridium stercorarium.* Murashima et al. mixed cellulosomal XynA, a hemicellulase, and other cellulases and incorporated them on truncated parts of scaffoldin protein CbpA of *Clostridium cellulovorans*. These minicellulosomes were found to degrade corn cell walls synergistically (Murashima et al., 2003). Thereby synergistic effects occurred not only if the hydrolases are bound to a protein carrier such as CipA or CbpA, but also if certain free hydrolases are mixed with other free hydrolytic enzymes. Zverlov and co-workers (2006) demonstrated that the simultaneous presence of Cel9I, a non-cellulosomal processive endo-glucanase, and Cel48Y, a non-cellulosomal glycosyl hydrolase, leads to a 2.1-fold higher activity on highly crystalline bacterial cellulose than was expected by additive activity of the single enzymes (Berger et al., 2006). Bronenmeier et al. (1999) combined the endoglucanase Cel9Z and the exo-1,4-β-glucanase Cel48Y from *Clostridium stercorarium*. The activitiy of the combined enzymes towards microcrystalline Avicel and BMCC was 2-3 times higher than the sum of the individual activities (Riedel et al., 1999, 1998).

5.2 Miniscaffoldin-enzyme complexes and CipA

Bayer et al. (2001, 2005) designed chimeric cellulosomes in which selected enzymes were incorporated in a specific location within a mulitcomponent complex. The chimeric scaffoldin was designed to include two cohesin modules from two different species, *C. thermocellum* and *C. cellulolyticum*, optionally connected to a cellulose binding module. Recombinant enzymes with corresponding dockerins from these bacterial bound only to their counterpart derived from the same host, not to that from the other (Fierobe et al., 2001). Thus, appropriate dockerin-containing enzymes could be assembled precisely and by design into the desired complex in which the components were arranged in a predetermined order and stoichiometry. Compared with the mixture of the free cellulases, the resultant cellulosome

chimeras exhibited enhanced synergistic action on crystalline cellulose (Fierobe et al., 2001, 2005).

This different but related strategy was used in this study to create the artificial cellulosome. Various components of the scaffoldin protein CipA were recombinantly produced and further investigated. Additionally, the complete CipA could be cloned and expressed successfully. Due to the presence of repetitive sequences within the *cipA* gene and the resulting difficulties for a successful cloning and expression procedure, many researchers had so far failed to obtain the complete purified CipA protein. With the usage of a specific expression system, the Champion pET-TOPO expression vector in combination with the expression in the Rosetta-Gami B *E. coli* strain, it was possible to produce the complete CipA protein. Thus, with the combination of the exo-enzymes of the SM901 mutant, an artificial cellulosome could be reconstituted *in vitro* for the first time. Initially, the created complex showed a diminished hydrolytic activity on microcrystalline cellulose (70 %) compared to a purified cellulosome preparation, but a 13 fold higher activity than the free unbound enzymes of the mutant. The diminished performance compared to the natural archetype led back to missing components in the mutant cocktail, which seem to be essential for effective complete depolymerization of crystalline cellulose. It could be further assumed that not all hydrolytic components are bound to the recombinant CipA during the experiments and therefore the hydrolytic performance is diminished compared to the theoretical power of the system or that the reconstitution was not perfect. The recombinant type of CipA, which is not glycosylated, may also have some effect.

Gel exclusion chromatography experiments confirmed the ability of recombinant CipA protein to bear at least seven dockerin bearing hydrolytic enzymes. This sample resulted in two main elution peaks, firstly with a molecular mass higher than 443 kDa representing CipA with interacted enzymes, secondly a wide peak in a range of 150 to 60 kDa, representing unbound SM901 mutant enzymes (Fig. 17).

However, the scaffoldin protein with its high molecular mass of ~200 kDa, full of repetitive sequences, was not produced by host cells in sufficient

amount. Thus, later commercial production in large scale will not be cost effective. Therefore, another strategy was pursued.

Truncated parts of the scaffoldin unit, called miniscaffoldins, were produced recombinantly and immobilized on the surface of magnetic nanoparticles (Fig. 51). These hydrolase carrier connected nanoparticles were loaded with the dockerin bearing hydrolases of the SM901 mutant cocktail to obtain a system functionally identical to the native cellulosome. To guarantee some flexibility of the nanoparticle miniscaffoldin system, a heterobifunctional PEG based linker between the miniscaffoldin and nanoparticle surface was interconnected. Activity assays with spacer bearing particles and particles bearing miniscaffoldins bound directly on the surface demonstrated that the interconnection of a linker results in an improved hydrolytic performance.

The single miniscaffoldins (Fig. 21) consist of different amount of cohesins and optionally a carbohydrate binding module (CBM) from *C. thermocellum*. They were tested for their ability to interact with dockerin bearing enzymes by native-PAGE. The analyses could confirm the correct function for cohesin-dockerin recognition (Fig. 25). Different miniscaffoldins (single cohesin, two cohesins, three cohesins, optionally each with a CBM) connected with nanoparticles were loaded with hydrolytic dockerin bearing exoenzymes from the SM901 mutant. Activity assays on soluble cellulose such as barley β-glucan and carboxy-methyl cellulose (CMC) showed no difference in hydrolysing these kinds of substrate just as with the free enzymatic components. The incorporation of the enzymes on miniscaffoldins plays no role if the substrate is soluble and non-crystalline.

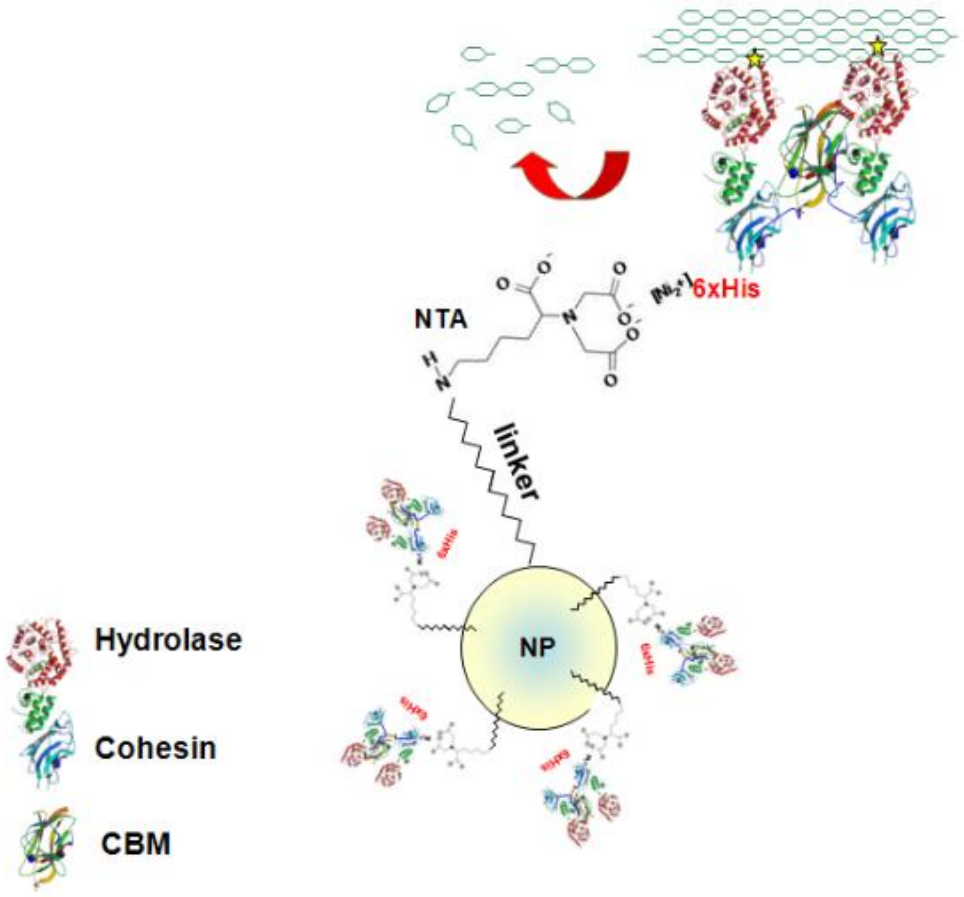


Fig. 51: Schematic representation of the artificial cellulosome. Hydrolases bearing miniscaffoldins (coh2-CBM-coh3) are bound on the surface of magnetic nanoparticles (NP) via an interconnected linker (protein structures courtesy of Carvalho et al., 2007).

However, dramatic differences could be observed with insoluble crystalline substrates such as MN300 or Avicel, in those complexes rearranged on nanoparticles bearing miniscaffoldins consisting of two or more cohesins. Nanoparticles bearing miniscaffoldins consisting of just a single cohesin did not show such an advance in activity towards crystalline substrates. With an increasing number of cohesins in the miniscaffoldins, an enhancement of hydrolytic action could be obtained in accordance with the cellulosome hypothesis. It was not surprising that a single immobilized cohesin showed no improvement in hydrolytic performance compared to the free enzymes. Indeed the immobilized cohesins are in close proximity, but the calculated distance on the nanoparticles of ~ 14 nm from one cohesin to another seems to keep two enzymes too separated to obtain synergistic effects between the cohesin bound hydrolases – the distance between two cohesins in CipA is about 4 nm. An enzyme bearing miniscaffoldin consisting of two cohesins exhibited a 1.7-fold increased degradation rate compared to the free

hydrolases. This could be explained by close proximity of the hydrolytic enzymes and the resulting synergism. This hypothesis is corroborated by the result from the miniscaffoldins containing only one cohesin which showed little to no improvement of activity. This strengthens the necessity for a certain distance between enzyme components within narrow limits in order to achieve optimum activity.

With the presence of a CBM in the miniscaffoldin, a further improvement of hydrolytic activity was achieved. CBMs seem to be in some way involved in the hydrolysis process (Din et al., 1991). They are believed to be important in increasing the local concentration of the catalytic moduls on the substrate and/or in disrupting the hydrogen bonds between cellulose chains in the crystall. These moduls play also important roles in free-acting enzymes (Tomme et al., 1988; Van Tilbeurgh et al., 1985). They belong to at least four different families, and a particular type of CBM is usually associated with a modul of a particular family (Ohmiya et al., 1997; Tomme et al., 1995): CBMs of family III of Cel9F and Cbh9A as well as CBMs of family IV of Cel9K and Cbh9A with family 9 catalytic moduls; CBM of family VII of Cel5E with a family 5 catalytic module; and CBM of family VI of Xyn10Z with a family 10 catalytic module (Ohmiya et al., 1997). The role of CBMs found in some cellulosomal catalytic components is not yet clear. The cellulosomal enzymes are attached to the cellulose surface by means of CipA containing a CBM of family III and in fact do not need their own CBMs.

It has been demonstrated that the family III CBM of CipA binds to both amorphous and crystalline cellulose, but its binding capacity with amorphous cellulose is 20 times higher (Morag et al., 1995). However, Cel9K CBM binds efficiently to acid-swollen cellulose and weakly to Avicel (Kataeva et al., 1998). The presence of different CBMs in several cellulosomal enzymes suggests that these modules play significant and specific roles with enzymes involved in cellulose degradation. The chemical simplicity of cellulose belies its structural complexity. The diversity of CBMs found in cellulosomal subunits may be necessary for the binding of the complex to various regions of cellulose, regardless of the degree of its crystallinity and other peculiarities of its structure.

With the introduction of a CBM III, which interacts with the cellulose fibers, in miniscaffoldin, the whole complex is able to get in close contact with the substrate. Thus a rapid interaction of hydrolases with the target is guaranteed. This effect is most prominent, if a miniscaffoldin bears a CBM in combination with three cohesins. A further improvement (3.5-fold) of the overall activity could be obtained with that kind of miniscaffoldin. Therefore, these results are in agreement with the cellulosome paradigm (Bayer et al., 1998). Near proximity of single hydrolytic components to each other and the high local concentration result in a high efficiency in the dismantling of crystalline cellulose. Similar results could be obtained from Fierobe et al. (2005). They designed trifunctional miniscaffoldins (3 cohesins from different species connected with a CBM) that were found to be considerably more active than a bifunctional system (2 cohesins from different species connected with a CBM) (Fierobe et al., (2005). Hence a higher amount of neighbouring cohesins in the miniscaffoldin leads to a higher hydrolytic action of the enzymes. Additionally, due to the presence of a CBM, the whole complex stands in direct contact to substrate. Thus, diffusion limiting factors are futile. Some experiments revealed that that the position of the CBM within the miniscaffoldin, between two cohesins or on the edge (CBM-cihesin3-cohesin4 and cohesin2-CBM-cohesin3), does not seem to play any role in the hydrolytic activity of the complexes.

The results of the hydrolytic activity on crystalline cellulose of the CipA protein and the miniscaffoldins loaded with SM901 exoenzymes and the comparison with native cellulosome indicates that some components are missing or underrepresented in SM901 enzyme cocktail. The mutant has to be grown on cellobiose; it is hardly propagated on Avicel. A different distribution of hydrolytic components within the cellulosome (actually, it is not present in the mutant) on these two substrates has been described (Bayer et al., 1985; Lamed et al., 1985; Freier et al., 1988; Morag et al., 1990; Blumer-Schuette et al., 2008; Gold et al., 2008). This may result in an infavourable composition of cellulosomal components in the culture supernatant compared to on crystalline cellulose grown cultures. Hence, in accordance with the studies of Gold et al. (2008) and Zverlov et al. (2004), additional hydrolytic major components of the cellulosome were produced

which were found to be underrepresented when grown on cellobiose and were added to the enzyme cocktail in different combinations and amounts to fill this gap.

5.3 Additional enzymatic components and their relative abundance induced by Avicel and cellobiose

In previous studies Bayer et al. (1985) investigated the cellulase system from *C. thermocellum* wild type and an adherence defective mutant. Specifically the growth conditions - growth either on crystalline cellulose (Avicel) or on cellobiose as insoluble or soluble carbon sources respectively - were found to be critical to the distribution of the cellulosomal proteins in the mutant system: the cellobiose-grown mutant (in contrast to the wild type) lacked the cellulosome on its surface and produced only minor quantities of the extracellular cellulosome accompanied by other relatively low-molecular-weight cellulases. The polypeptide composition of the respective purified cellulosome was dependent on the nature of the carbon source (Bayer at al., 1985). Later this knowledge was verified and manifested by Gold et al. (2008) by gene expression analysis (Fig. 52).

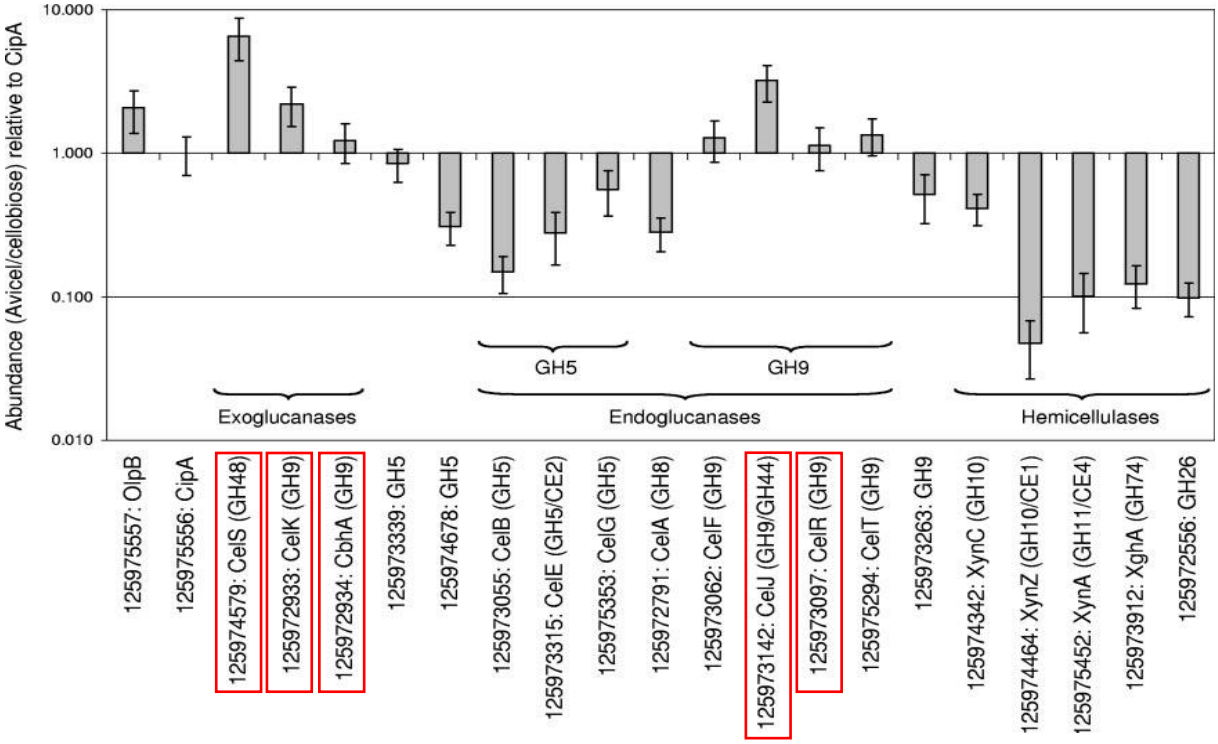


Fig. 52: Fractional differences in expression of *C. thermocellum* Avicel-grown cellulosomal components relative to cellobiose-grown components, normalized to CipA, over a logarithmic scale (Gold et al., 2008). The recombinantly produced cellulases in this study are marked red.

The exoglucanase Cel48S exhibited the greatest increase of any docking component during growth on Avicel compared to cellobiose (Gold et al., 2008). The increase of Cel48S on Avicel versus cellobiose had already been observed at the protein level by SDS-PAGE (Bayer et al., 1985) and Western blot analysis (Dror et al., 2003). For effective dismantling of crystalline cellulose, exoglucanases are the key enzymes, so it was not surprising that exoglucanase Cel9K and Cbh9A also showed increased expression level on Avicel (Gold et al., 2008). Docking proteins with known endoglucanase activity demonstrated varied expression patterns. The GH5 endoglucanases Cel5B, Cel5E, and Cel5G demonstrated higher expression when cells were grown on cellobiose than on Avicel. The same was true for Cel8A from GH8. In contrast, Cel9J and Cel9R from GH9 showed increased expression on Avicel, while the expression of other GH9 endoglucanases, Cel9F and Cel9T, did not change significantly (Gold et al., 2008).

These data are in accordance with the previous studies of Zverlov et al. (2004). Here, cellulosomes were purified and the components were separated by an adapted two-dimensional gel electrophoresis technique. The apparent major spots were identified by MALDI-TOF/TOF: the structural protein CipA, the endo-glucanases Cel8A, Cel5G, Cel9N, Cel9R, the cellobiohydrolases Cbh9A, Cel9K, Cel48S, the xylanases Xyn10C, Xyn10Z, Xyn10D, the xyloglucanase Xgh74A, and the chitinase Chi18A (Zverlov et al., 2004, Zverlov et al., 2008).

According to these studies the chosen additional components seemed to be a promising approach in order to enhance the hydrolytic activity of SM901 mutant enzymes on crystalline cellulose. The different cellulases could be successfully expressed in *E. coli*, and their behaviour on soluble, amorphous and crystalline cellulose was determined.

The expression of Cel48S resulted initially in the formation of inclusion bodies (Wang et al., 1994). These issues often are a result of improper folding of the expressed proteins. Cel48S was purified under denaturing conditions with 5M urea and subsequent renaturation by dialysis. The resulting protein was active against soluble and insoluble cellulose. Unfortunately however, the essential amount of refolded and unfolded protein respectively could not

be testified and probably the activity of the native protein could not be achieved.

The recombinantly produced cellobiohydrolase Cbh9A showed atypical electrophoretic mobility in SDS-Page. The calculated mass of Cbh9A was determined to be 138 kDa. The actual, purified enzyme exhibited a molecular mass of ~ 100 kDa in SDS-gel. The same result occurred in the studies of Koeck (2008); exo-proteins of mixed cultures from a fermenter containing amongst other *C. thermocellum* cells, consequently also the exoglucanase Cbh9A, were separated by SDS-PAGE. SDS band of Cbh9A with a size of ~ 110 kDa was excised and N-terminal amino acid sequence could be confirmed by MALDI-TOF.

The regions of the catalytic modules of Cbh9A and Cerl9K have identity of about 90% on both nucleotide and amino acid levels, a level of homology that has not been found between any other two genes of the *C. thermocellum* exoenzymes sequenced to date (Béguin et al. 1996). It has been proposed that catalytic sites of *celK* and *cbhA* arose from a common ancestral gene by duplication. The fact that the catalytic moduls of Cel9K and Cbh9A were duplicated and then favourably selected suggests that these catalytic moduls may be essential for the bacterium to engage in cellulose degradation (Kataeva et al., 1999).

For a long time it has been believed that the cellulosome contains mostly endoglucanases (Mayer et al., 1984). The discoveries of CelS (Kruus et al., 1995), Cbh9A (Zverlov et al., 1998), and Cel9K (Kataeva et al., 1999) indicate that cellobiohydrolases play an important role in cellulose degradation by the cellulosome of *C. thermocellum*. This idea is further supported by the fact that Cel48S and Cel9K are the most abundant components of the cellulosome. Exoglucanases are in general the most abundant proteins, even in the fungal cellulase systems – also in *Clostridium cellulovorans* (Schwarz, personal communication).

Besides the non-catalytic structural OlpB protein and scaffoldin protein CipA, the processive endoglucanase Cel9J is the largest single component of the cellulosome. Cel9J is a modular enzyme composed of an N-terminal signal peptide and six modules in the following order: a carbohydrate binding module (CBM), an immunoglobulin like fold, a subfamily E1 (GH9)

endoglucanase module, a family J (GH44) endoglucanase module, a docking module, and another module of unknown function. Cel8A and Cel48S contain family D and L catalytic modules, respectively. Cel9J contains subfamily E1 and family J modules. Since the strong hydrolytic activity toward crystalline cellulose of the cellulosome of *C. thermocellum* is believed to be due to the synergistic action of individual enzymes, the presence of two or more catalytic modules from different families seems to be necessary to hydrolyze the substrate in a cooperative fashion. Therefore, Cel9J, which has two catalytic modules different from those contained in Cel8A and Cel48S and is a major protein in the cellulosome, must play an important role in the efficient hydrolysis of crystalline cellulose (Ahsan et al., 1996). Activity assays by Arai et al. (2002) showed that CBM30 of Cel9J is extremely important for its activity, not only because it mediates the binding of the enzyme to the substrates, but also because it participates in the catalytic function of the enzyme or contributes to maintaining the correct tertiary structure of the family 9 catalytic module for expressing enzyme activity (Arai et al., 2002).

Cel9R, another major component in the cellulosome of *C. thermocellum*, is one of the most prevalent β-glucanases in the complex after Cel48S and Cel8A (Zverlov et al., 2005). Cel9R as well as Cel9J behaves as a processive endoglucanase. A processive enzyme would exclusively produce short cellodextrins which initially appear in the soluble fraction, whereas an endo-glucanase should – at least initially – produce new reducing ends exclusively in the insoluble fraction. Cel9R produced almost equal amounts of reducing residues in the soluble as well as in the insoluble phase (Zverlov et al., 2005), whereas reducing ends were only found in soluble fraction in case of hydrolysis with non-processive endoglucanases, e.g. Cel8A (Schwarz et al., 1986).

In this work the three mentioned exoglucanases (Cbh9A, Cel9K, and Cel48S) and the two endoglucanases (Cel9J, Cel9R) were produced recombinantly and added to the enzyme-cocktail of the *Clostridium thermocellum* SM901 mutant. These hydrolases answer the purpose of restocking missing cellulosomal components of the cellobiose adapted SM901 mutant cells.

5.4 Addition of main cellulosomal components to SM901 mutant cocktail

In previous works, single enzyme components of the *C. thermocellum* cellulosome have been shown individually to exhibit enhanced activity on insoluble cellulose substrates upon incorporation via a suitable scaffoldin into a cellulosome-like complex. In an early study, Wu et al. (1988) reported that a purified cellulosomal cellulase (Cel48S) can be combined with the native scaffoldin, leading to an increase in hydrolytic activity of the complex on crystalline cellulose. More recently, Kataeva et al. (1997) showed that a different cellulosomal enzyme (endoglucanase Cel9D) interacts stoichiometrically with scaffoldin constructs, and the resultant complexes were found to degrade cellulose in a synergistic manner. Yet another cellulosomal enzyme (endoglucanase Cel5E) was shown by Ciruela et al. (1998) to exhibit enhanced crystalline cellulase activity upon prior interaction with the full-length recombinant scaffoldin. In each of these latter studies, only one enzyme type was incorporated into the given complexes, and the observed enhancement of activity was attributed mainly to targeting of the enzyme to the solid substrate by the scaffoldin-borne CBM. Finally, Bhat and colleagues reconstituted a simplified cellulosome by combining purified preparations of native cellulosomal components, including the full-size scaffoldin with selected enzymatic subunits. The resultant reconstituted complex exhibited enhanced synergy on cellulose compared with the activity of the mixture of free enzymes.

In this study the cellulosomal major components (Cbh9A, Cel9J, Cel9K, Cel9R, Cel48S, relating to Avicel grown culture) were recombinantly produced, mixed with SM901 exoenzymes and bound to CipA protein to optimize the enzyme composition, obtained from an on cellobiose grown SM901 mutant of *C. thermocellum*, to degrade crystalline cellulose. Their positive influence towards the hydrolytic performance of the SM901 enzyme cocktail was further investigated.

The exo- and endoglucanases were mixed with SM901 mutant enzymes in different amounts and combinations. Initially just the single components were bound to CipA. In comparison to free unbound state all hydrolases combined with CipA showed 8 to 9 fold increased hydrolytic activity towards

Avicel. Due to the close proximity limiting diffusion effects were diminished and due to the targeting effect to the substrate caused by the CipA internal CBM, a dramatic enhancement of hydrolytic action could be observed.

Next, single recombinant components were mixed with mutant enzymes in a 1:10 ratio and bound to CipA. The aim was to identify missing components of enzyme cocktail from on cellobiose-grown mutants. No explicit improvement in comparison to mutant enzymes alone could be achieved. This indicates that the addition of at least two or more components to mutant enzymes would be necessary to compete with the performance of purified cellulosome. Different samples with the endo- and exoglucanases and the mutant enzymes lead to a combination similarly effective as the purified cellulosome: the stoichiometric mixture of 10% Cel9J, Cel9K, Cel9R, Cel48S respectively and 60% enzyme cocktail bound on scaffoldin protein CipA. Further experiments should be carried out to answer the question whether the shifting of the amount of the single components in the mixture could achieve further improvement. Fierobe et al. (2005) showed that the different stoichiometric distribution of the type of the single enzymatic components bound on a miniscaffoldin could play an important role in degradation of crystalline cellulose. Samples with two molecules of the endoglucanase Cel9G and one molecule of the endoprocessive cellulase Cel48F from *Clostridium cellulolyticum* per miniscaffoldin (three cohesins and a CBM) showed a ~ 1.5 fold enhancement of specific activity compared to samples containing just one Cel9G and two Cel48F per miniscaffoldin (Fierobe et al., 2005). Studies from Zverlov and co-workers reported that the enzyme Cel48Y shows a distinct synergism of 2.1 times with the noncellulosomal processive endoglucanase Cel9I (both from *C. thermocellum*) on highly crystalline bacterial cellulose at a 17-fold excess of Cel48Y over Cel9I (Berger et al., 2006).

It seems not surprising that Cel48S, the main cellulosomal component, appeared in this mixture. Exoglucanases are the key enzymes in cellulase mixtures effective on crystalline cellulose (Teeri et al., 1997). Another exoglucanase, the cellobiohydrolase Cel9K took part in this mixture. Previous studies indicate that these two cellobiohydrolases (Cel48S and Cel9K) play an enormous role in cellulose degradation by the cellulosome of

C. thermocellum (Kruus et al., 1995; Kataeva et al., 1999). In addition to these two exoglucanases, the two endoglucanases CelJ and Cel9R were part of the combination. Single Cel9J, the largest catalytic subunit of the cellulosome, showed an almost equal catalytic activity towards Avicel as Cel48S. It suggests that a catalytic subunit other than Cel48S capable of hydrolyzing crystalline cellulose plays an important role at least in the early stage of the hydrolysis of crystalline cellulose, though the activity of Cel9J is much lower than that of the cellulosome (Ahsan et al., 1997).

CelR, which is the most abundant endoglucanase in cellulosomes, is one such enzyme, a processive GH9 endoglucanase that produces cellotetraose as its primary hydrolysis product (Zverlov et al., 2005). These kinds of cellodextrins could be distinguished substrates for exoglucanases, like Cel9K and Cel48S, which gives another hint of a co-working process between endo- and exoglucanases resulting in synergistic action.

The cooperation of the two different types of glucanases, the exo- and endoglucanase, plays an enormous role in the degradation of crystalline cellulose. Endoglucanases unhinge single strands of the crystal structure and points of action for exoglucanases are thus provided. Due to their processive mode of action the exoglucanases migrate alongside unhinged strands releasing mainly cellobiose as hydrolysis products. Because of their close proximity to each other in formed miniscaffoldin-complexes their synergistic action could be enhanced. Thus diffusion-limiting steps are negligible and the search of new attack sites for exoglucanases, and their resultant time costing “search” for other sites is diminished through permanent preparation of new points of action by the closely located endoglucanases.

The fifth recombinant component Cbh9A seemed not to be playing a relevant role in the tested reactions. Cbh9A has a multidomain structure; among other things it consists of an N-terminal and an additional C-terminal CBM (family III and IV). It should be kept in mind that family IV and family III CBMs differ strikingly in their substrate specificity (Zverlov et al., 1998). Whereas family III moduls bind specifically to crystalline cellulose, family IV moduls bind with approximately equal affinities to amorphous cellulose, cellooligopentaose, and mixed-linkage β-glucans (Johnson et al., 1996;

Tomme at al., 1996). Maybe, due to CBM moduls in Cbh9A, it is not as necessary to be supported by the CBM present in miniscaffoldin or CipA as for CBM missing hydrolases. Whereas, due to lack of CBMs, CelS for example, requires the presence of CipA or at least a CBM bearing miniscaffoldin for the efficient hydrolysis of crystalline cellulose. Thus, the diminished contribution of Cbh9A in the mixtures could be explained.

The performance of the mixture, containing CipA protein with SM901 mutant enzymes and the mentioned exo- and endoglucanases which displayed the best performance in degradation of crystalline cellulose, was compared with a newly generated commercial enzyme preparation (Biopract GmbH) (Fig. 53). The commercial product is based on the fungal enzyme system of *Trichoderma reesei* and specially offered for the degradation of crystalline cellulose. The enzyme complexes (except for the commercial product) were incubated with the dockerin fused β-glucosidase from *Thermotoga neapolitana* to avoid product inhibition by the main product cellobiose and to get a comparable result to this of fungal cellulase based product, which contain at least one enzyme capable for cellobiose degradation.

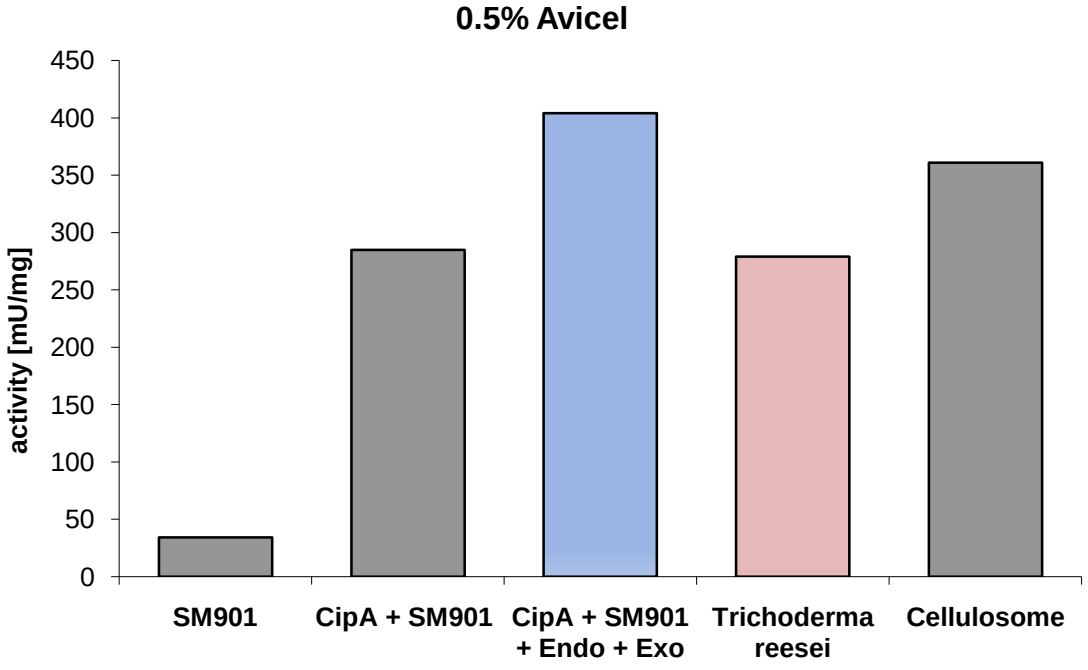


Fig. 53: Specific hydrolytic activities [mU/mg] of the mixture (CipA, SM901 mutant enzymes, Cel9J, Cel9K, Cel9R, Cel48S [blue]) compared with a commercial enzyme preparation of Biopract GmbH (*Trichoderma reesei* [red]).

Thin-layer-chromatography of the degradation products by fungal cellulases resulted mainly in glucose and an activity assay with p-nitrophenyl-β-D-glucoside indicated that a β-glucosidase was already present in the cellulase mixture. Figure shows that the chosen mixture exhibited a higher degradation rate (~35 %) than the commercial product and even higher (~13 %) than the purified cellulosome. The reaction mixtures with the commercial product took place under the supplier's recommended conditions (pH 5, 50 °C), the developed mixture at pH 6.5 and 60 °C. Due to the additional incorporation of the modified β-glucosidase in the nanoparticle-miniscaffoldin-enzmye complex, product inhibition by the main product cellobiose could be avoided and the degradation of crystalline cellulose yielded mainly in the fermentable product glucose. Furthermore, the produced hydrolytic complex exhibited a higher hydrolytic activity than a native cellulosome preparation. Previous studies showed that specific hydrolysis rates of metabolically active cultures of *C. thermocellum* displaying cellulosomes are more than fourfold higher than those of purified cellulosomes (Lu et al., 2006). However, this factor may differ with the type of substrate and should always be determined in direct comparison and under conditions optimal for both enzyme systems.

5.5 Impact of β-glucosidase BglB from *Thermotoga neapolitana* fused with *C. thermocellum* dockerin type I

Hydrolysis of crystalline cellulose by SM901 mutant enzymes and also by recombinant cellulases resulted mainly in cellobiose as the endproduct (Fig.). It is reported that cellobiose could act as a competitive inhibitor for cellulases (Kadam et al., 1989; Halliwell et al., 1973; Katz e al., 1968; Howel et al., 1975; Reese et al., 1952). The degree of inhibition of cellobiohydrolases by cellobiose depended on the relative concentrations of substrate and inhibitor. Increasing the relative abundance of substrate over the product enabled the inhibition to be gradually relieved and almost overcome (Halliwell et al., 1973). To prevent this problem, the extremely thermostable β-glucosidase BglB from *Thermotoga neapolitana* was introduced to the enzymatic complex. In order to enable later coupling on immobilized miniscaffoldins the enzyme was fused C-terminally with the dockerin type I

of the *C. thermocellum* endoglucanase Cel8A. For this a dockerin-fusion vector based on the expression vector pQE (Qiagen) was constructed. This vector can be used for dockerin fusion of all proteins of interest from other organisms than *C. thermocellum* to introduce them in miniscaffoldin-nanoparticle complex. Therefore, an extension of the complex with e.g. fungal hydrolytic enzymes or enzymes from other cellulolytic active organisms should be possible.

The obtained β-glucosidase BglB*-Doc showed maximum activity at a temperature of 90 °C and at a pH of 6 (Zverlov et al., 1997). Hydrolytic activity at 60 °C reached 70 % of the maximum. The ability of a fused dockerin to interact with cohesin containing miniscaffoldins could be verified by native PAGE analysis. The mixture of the enzyme and miniscaffoldin resulted in a single band with altered mobility, indicating that complete or near complete complexation could be achieved.

The performance of β-glucosidase mixed with components of SM901 mutant enzymes was determined by activity assays on crystalline cellulose. The inhibitory behaviour of the major endproduct cellobiose could be diminished by splitting cellobiose into glucose monomers. The DNSA-reducing power in the samples with added β-glucosidase is twice as much as without the enzyme BglB. This indicates that cellobiose was degraded by the enzyme to glucose, resulting in nearly twice the number of detectable reducing ends. In contrast to SM901 mutant enzymes (~ 80h), the saturation (inhibition) of the curve alone, and consequently the stagnation of hydrolytic action, could be decelerated to almost 140 hours. It is reported that glucose could also act as an inhibitor even if not with the same impact as cellobiose. In fungal systems, increasing glucose content in the hydrolysate dramatically increased in the degree of inhibition of both β-glucosidase and cellulase activities (Xiao et al., 2004). In which way this knowledge is transferable to the enzymatic system of *Clostridium thermocellum* has not been reported so far. For future application, processes for the removal of the accumulated glucose may have to be developed.

5.6 Xylanase-miniscaffoldin complexes

Even if cellulases are optimally expressed by growing the bacterium on cellulose, about half of the proteins in the cellulosomes are hemicellulolytic enzymes (Zverlov et al., 2005). Of the 71 genes potentially encoding cellulosomal hydrolytic subunits in *Clostridium thermocellum*, 17 code for non-glucanolytic poly-saccharide hydrolases (including xylanases), 8 for glycosidases and 4 for esterases (Zverlov et al., 2005). Thus a large fraction of the hydrolytic genes is not directly related to cellulose degradation. The function of the non-cellulolytic enzymes is presumably the unwrapping of the cellulose crystals from the covering matrix of lignin, pectin and hemicellulose (Zverlov et al., 2005).

Isolated xylans, the major part of hemicellulose, are typically polydispersed hetero-polysaccharides and comprise a backbone of β-1,4-linked D-xylopyranosyl residues. The xylopyranosyl backbone is substituted at positions C-2, C-3, C-5 to varying degrees depending upon the plant and the stage of development of the plant when the polymer was obtained (Joseleau et al., 1992; Wong et al., 1988). In monocots, at the C-2 position 1-2-linked α-D-glucuronic acid or 4-*0*-methyl-α-D-glucuronic acid might occur, while at C-3 of xylopyranose, one frequently finds 1-3-linked α-L-arabinofuranose. In some xylans, particularly in hardwoods, xylopyranose may be *0*-acetylated at the C-2 or (more commonly) the C-3 position. Additionally, a small, but important amount of phenolic components, such as ferulic and *p*-coumaric acids (associated with lignin), may be esterified to xylan via their carboxyl groups to C-5 of arabinose branches (Kato et al., 1985).

As a first attempt to degrade the hemicellulose together with cellulose, Murashima and co-workers (2003) created minicellulosomes consisting of truncated parts of scaffoldin protein CbpA of *Clostridium cellulovorans*, a hemicellulase (XynA) and three cellulases (endoglucanase EngE, EngH and exoglucanase EngS). The minicellulosome showed high synergistic action on the natural substrate corn cell wall (Murashima et al., 2003). This emphasizes the importance of the hemicellulases, together with cellulases, in the degradation of natural cellulosic biomass and indicates that an introduction of at least one hemicellulose in a miniscaffoldin-cellulase complex is essential for effective hydrolytic depolymerization of natural substrates. In this study no hemicellulases together with cellulase were

introduced in miniscaffoldins. However, a potential synergistic effect between two xylanases incorporated on a miniscaffoldin was further investigated.

Some dockerin bearing xylanases of *C. thermocellum* (Xyn11A, Xyn10C and Xyn10Z) and a xyloglucanase (Xgh74A) were tested for synergistic behaviour by complexing with CBM-bearing miniscaffoldin containing two cohesins. Whereas synergistic action between cellulases was previously reported in numerous publications, these effects were investigated in this work for the first time. Mixtures of two differerent xylanases were prepared and incubated together with minisacffoldin (CBM-cohesin3-cohesin4) on purified oat spelt xylan. Xyloglucanase Xgh74A showed no activity on both insoluble and soluble fractions of oat spelt xylan, neither in complexed form nor in a free state as could be expected. Oat spelt xylan consists mainly of arabinoxylan and in low parts of glucorono-arabinoxylan. This kind of substrate exhibits no xyloglucan structures that are hydrolytically accessible for Xgh74A. Otherwise strong hydrolysis could be detected with soluble cellulose barley β-glucan. The high activity of Xgh74A on mixed-linkage barley β-glucan is not unusual among xyloglucanases. The *Thermotoga maritima* xyloglucanase Cel74 has its highest activity on barley β-glucan (Chhabra and Kelly, 2002) and Xgh74A shows 38 % sequence identity in the catalytic module (Zverlov et al., 2005). The removal of the xyloglucan in plant cell walls is obviously a precondition for efficient cellulose hydrolysis, and Xgh74A may be able to perform this task in *C. thermocellum*. Therefore, the xyloglucanase XghA could play an important role in the nanoparticle-miniscaffoldin-enzyme-complex for effective degradation of natural substrates.

Enhanced activity of the other produced xylanases towards insoluble oat spelt xylan fraction of the xylanase-miniscaffoldin complexes due to synergistic effects could be detected. The molecular structure of xylan could be changed by the heating process for fractionizing the xylan, however, into soluble and insoluble components (Schwarz and Zverlov, personal communication).

The combination Xyn10C and Xyn10Z with miniscaffoldin should be distinguished, as an enhanced activity of about 27.5% was notified. The hydrolytic performance of Xyn10Z could be impaired by complexation. This xylanase showed the strongest hydrolytic activity by immobilization of the

enzyme alone. Grépinet et al. (1988) reported that XynZ is highly active on the chromogenic and fluorogenic substrates such as pNP-β-D-xylobioside, pNP-β-D-xyloside, pNP-β-D-glucoside, pNP-β-D-cellobioside, MU-β-D-cellobioside and MU-α-L-arabinoside. Thus, cleavage of the heterosidic bond appears to be less specific than cleavage of the holosidic bond, since xylanase XynZ is inactive towards CMC and cellodextrins (Grépinet et al., 1988). The α-L-arabinose side chains of the oat spelt xylan were thus removed by XynZ and the blank xylose chain is more accessible for effective depolymerization.

Although Xyn10C and Xyn10Z both belong to the family GH 10, both enzymes complement each other well and apparently exhibit together a more efficient hydrolysis of xylan. The two different binding moduls, CBM22 (Xyn10C) and CBM6 (Xyn10Z), probably guarantee a particularly advantageous localization of the catalytic moduls on the substrate. Thus the substrate concentration is increased around the enzymes and a much easier degradation of the substrate is possible.

5.7 Coupling on nanoparticles

In this study nanoparticles were used to immobilize hydrolase bearing miniscaffoldins. Miniscaffoldins were interconnected between the beads and the hydrolases. This guarantees cohesin-dockerin interaction without the loss of activity caused by immobilization. Through the presence of a carbohydrate binding module in miniscaffoldins, the whole complex could come into close contact with substrate. The nanoparticle surface with amine or carboxy functional groups was covalently crosslinked to a heterobifunctional PEG-based linker and finally to nitrilo-triacetic acid (NTA) by glutaraldehyde (amine) or EDC/Sulfo-NHS (carboxy) respectively. Nickel ions for interaction with 6xHis tagged miniscaffoldins were chelated by NTA molecules immobilized on the nanoparticle surface. The resulting nanoparticles reserve two binding sites for the interaction with the poly-histidine structure of the fusion peptide, the miniscaffoldins, where two vacant coordination sites on the nickel ions are exposed and enable strong coordination with the electron-donor containing pouch in the poly-histidine

sequences (Shieh et al., 2006). The best coupling results were obtained with carboxy modified particles with the application of 2 mM EDC and 5mM Sulfo-NHS. Thus, a coupling efficiency of about 80 µg miniscaffoldin protein/mg particle could be achieved. In contrast to amino-modified beads, the best efficiency was diminished to a value of about 62 µg miniscaffoldin protein/mg particles. The developed nanoparticle-miniscaffoldin system is reversible. Due to the chelating effect of disodium ethylenediamine tetra-acetate (EDTA), poly-histidine-fused proteins could be removed from particles by treatment with EDTA (van Ketel and Bruynzeel, 1984). After subsequent nickel removal and rechargement, particles were able to bind ~ 50 µg (62.5 % recovery) of miniscaffoldin protein, and twice recycled nanoparticles could immobilize ~ 25 µg (31.2% recovery) minscaffoldin protein.

Activity assays clearly demonstrated that the hydrolytic performance of immobilized enzymes is not negatively affected by the immobilization as is often the case with other methods. Due to their interaction with interconnected miniscaffoldins, the enzymes stay in their natural conditions. It is unclear whether all immobilized enzymes are able to stay in contact with substrate. Because of the spherical morphology of nanoparticles it could be possible that simply one enzyme-bearing hemisphere turned to substrate is active and the turned away site is not. However, by interconnection of a heterobifunctional PEG based linker a kind of flexibility in the system is granted. Samples with and without interconnection of a linker showed that hydrolytic action towards crystalline cellulose could be slightly improved. The spatial flexibility added by the linker was thus an important factor for activities on the insoluble substrat. An important point for future application is the enhanced stability of the nanoparticle-enzyme complexes. Kinetic and stability studies showed that the enzyme activity was better preserved upon binding onto the nanoparticles when subjected to thermal and various pH conditions (Kouassi et al., 2005).

Covalent binding provides the enzymes with protection against structural denaturation due to the unfavorable solvent-protein interactions, and results in activation effect, a possible reason for the better activity of the bound enzyme compared with the free enzyme after heat treatment and pH value

changes (Wang et al., 2001). Furthermore these observations can be confirmed by long term studies. Immobilized enzymes showed an almost unaltered stability over the time period in contrast to the free enzymes, which displayed a diminished activity of 40 % after 45 days at 60 °C. Due to their magnetic behavior the particles could be separated from the surrounding solution. The determined recovery rate by application of an external magnetic field was nearly 100 % (91 to 97 %). For future applications, this could be an important aspect for recycling enzymes from the reaction solution, although common hydrolytic preparations have to be refilled in the reactor after every reaction cycle.

Chapter 6

Summary

In our current time it is essential to aspire towards fossil fuel independent, alternative energy sources. Lignocellulose containing biomass, the most abundant single organic compound on earth, is an ideal raw material for chemical products or for the fermentation-process to biofuels. Efficient cellulase and hemicellulase degrading enzymes are thus a precondition for the production of 2nd gerneration biofuels. The hydrolytic system of the anaerobic thermophilic bacterium *Clostridium thermocellum* is a promising candidate as it is the most efficient organism producing hgydrolytic enzymes (or an enzyme system) known so far. *C. thermocellum* achieves efficient cellulose hydrolysis using multiprotein extracellular enzymatic complexes, termed cellulosomes. Its *in vitro* reconstitution was the goal of this thesis.

The cellulosome is a complex consisting of many different species of enzymes and a very special non-catalytic subunit, the scaffoldin, that plays multiple roles in cellulosme structure and function. The scaffoldin subunit mediates both the binding to cellulose via an internal cellulose binding module (CBM) and the attachement of many catalytic units via a set of internal closely related modules – the cohesins. The enzymatic subunits contain a catalytic module and a dockerin module. The dockerin module is able to interact with one of the cohesin on the scaffoldin subunit. Thus, the hydrolytic enzymes are incorporated onto the scaffoldin unit via the strong cohesin-dockerin interaction.

In this study an artificial cellulosome based on modular structure of the *Clostridium*-cellulosome was developed which exhibits at least the degradation level of the native cellulosome. Due to the combination of molecular biological-, peptide-chemical- and nano-technological-methods it was possible to realize the artificial reconstitution of the hydrolytic complex. In this system, hydrolytic cellulases from a *C. thermocellum* mutant are bound to the surface of nanoparticles.

To avoid negative affects due to the immobilization of the enzymatic components, recombinant miniscaffoldins, derivates from scaffoldin CipA protein, act as protein carrier and are connected by an inter-connected heterobifunctional PEG-based linker to the surface of the nanoparticles. Due to the super-paramagnetic behaviour of the nanoparticles, the enzymatic system could be recovered from reaction solution by an external magnetic field.

The nanoparticle-enzyme complexes showed enhanced hydrolytic activity towards high crystalline cellulose compared to the free unbound enzymes. Cellulosomal components overexpressed by growth on cellulose (in contrast to growth on cellobiose) were cloned and added to the mutant enzyme cocktail. By addition of the selected recombinant cellulases the hydrolytic activity could be enhanced further. Additionally, a thermostable β-glucosidase of *Thermotoga neapolitana* was fused with a dockerin and integrated into the nanoparticle-enzyme system. Thus, product inhibition by the main product cellobiose could be avoided and the degradation of crystalline cellulose yielded mainly in the fermentable product glucose. A composition of cellulases (selected exo- and endoglucanases, SM901 mutant enzymes and the β-glucosidase) together with the complete recombinant CipA protein exhibited an at least equal degradation level as the purified native cellulosome.

This composition can actually compete against a commercial available cellulose preparation, it exhibited even a ~30 % higher degradation level than the fungal based system under the optimal conditions respectively. Due to the poor expression level by the host *E. coli*, the recombinant CipA protein could not be purified in sufficient quantities for immobilization on the nanoparticles surface. The results with miniscaffoldins bound to the beads clearly showed however that the immobilization has had no negative effect on the hydrolytic performance.

Based on this aspect and the possibility to integrate the complete CipA protein into the nanoparticle-enzyme complex, in future a highly active renewable system could be developed, that is able to compete with the commercially available cellulase based preparations.

Zusammenfassung

In der heutigen Zeit besteht aufgrund der zunehmenden Knappheit an fossilen Brennstoffen und steigender Bedeutung des Klimaschutzes ein wachsendes Interesse, Energie aus eneuerbaren Rohstoffen zu gewinnen. Lignocellulose-haltige Biomasse ist die weltweit am häufigsten vorkommende organische Verbindung und stellt somit als großzügig verfügbarer umweltfreundlicher Rohstoff ein ideales Ausgangsprodukt für die Herstellung von Chemikalien und Biotreibstoffen dar. Effiziente Cellulose- und Hemicellulose-abbauende Enzyme sind deshalb für die Produktion von Biokraftstoffen der zweiten Generation eine Grundvoraussetzung. Das hydrolytische System des anaeroben, thermophilen Bakteriums *Clostridium thermocellum* ist hierfür ein vielversprechender Kandidat, da es als das effektivste der bisher bekannten hydrolytischen Enzyme oder Enzymsysteme gilt. *Clostridium thermocellum* betreibt effiziente Hydrolyse von Cellulose mit Hilfe eines extrazellulären, enzymatischen Multienzyme-Komplex, das Cellulosom. Ziel dieser Arbeit war die *in vitro* Rekonstruktion dieses Komplexes.

Das Cellulosom besteht aus verschiedenen Typen von Enzymen und einer sehr speziellen nicht-katalytischer Untereinheit, dem Scaffoldin-Protein (CipA), welches eine strukturelle und funktionale Rolle im Cellulosomen spielt. Die Scaffoldin-Einheit vermittelt einerseits die Bindung des Komplexes an das Substrat über ein Karbohydrat-Bindungs-Modul (CBM) und andererseits die Verankerung vieler katalytischer Komponenten über hochkoservierte, interne Module, Cohesine genannt. Die enzymatischen Untereinheiten bestehen wiederum neben ihrem katalytischen Modul aus einem Dockerin-Modul, welches mit einem Cohesin im Scaffoldin interagieren kann. Somit können die hydrolytischen Enzyme über die starke Cohesin-Dockerin-Interaktion in das Scaffoldin eingebaut werden.

In dieser Studie wurde ein künstliches Cellulosom basierend auf der beschriebenen modularen Struktur des *Clostridium*-Cellulosoms entwickelt, welches mindestens die hydrolytische Aktivität des nativen Cellulosoms aufwies. Durch die Kombination von molekular-biologischen, peptid-chemische und nano-technologischen Methodiken war eine erfolgreich

Zusammenstellung dieses artifiziellen Komplexes möglich. Hydrolytische Enzyme einer Scaffoldin-Mutante von *Clostridium thermocellum* wurden hierbei auf die Oberfläche von Nanopartikeln gebunden. Um negative Einflüsse durch die Immobilisierung der enzymatischen Komponenten zu vermeiden, wurden rekombinante Miniscaffoldine, Derivate des CipA-Proteins, als Protein-Träger für die Hydrolasen verwendet und mit Hilfe eines zwischengeschalteten heterobifunktionalen, auf PEG-basierenden, Linker auf die Oberfläche der Nanopartikel gebunden. Auf Grund des super-paramagnetischen Charakters der Partikel, kann dieses hydrolytische System aus den Reaktionslösungen mittels eines externen magnetischen Felds wieder gewonnen werden.

Der Nanopartikel-Enzym Komplex wies eine erhöhte hydrolytische Aktivität auf kristalliner Cellulose im Vergleich zu den ungebundenen Enzymen auf. Weitere Cellulosomale Komponenten, die im Organismus durch das Wachstum auf Cellulose überexprimiert werden (im Gegensatz zum Wachstum auf Cellobiose), wurden anschließend kloniert und dem Enzym-Cocktail der Mutante zugeführt.

Durch diese Zugabe konnte eine weitere Steigerung der Aktivität erzielt werden. Zusätzlich wurde eine thermostabile β-Glucosidase von *Thermotoga neapolitana* mit einem Dockerin-Modul fusioniert und ebenfalls in den Nanopartikel Komplex integriert. Auf diesem Weg konnte die, durch das hauptsächlich auftretende Endprodukt Cellobiose, bedingte Produkt-hemmung vermieden werden. Außerdem lieferte nun der Abbau der kristallinen Cellulose anstelle des Disaccharid Cellobiose überwiegend das fermentierbare Monosaccharid Glucose. Eine Kombination aus Mutanten-Enzymen, rekombinanten Komponenten und der β-Glucosidase zusammen mit dem CipA Protein erreichte nun mindestens das Abbaulevel eines nativen Cellulosoms aus *Clostridium thermocellum*.

Diese Zusammensetzung der Komponenten konnte sogar mit einer kommerziell erhältlichen Cellulase-Präparation, basierend auf Pilzenzymen konkurrieren, und erzielte eine um 30 % höhere Abbaueffizienz von Cellulose unter den jeweiligen optimalen Reaktionsbedingungen. Auf Grund des niedrigen Expressionslevel durch den Wirt, konnte das Scaffoldin Protein CipA nicht in ausreichenden Mengen für die Immobilisierung hergestellt

werden. Dessen ungeachtet zeigten die Versuchsergebnisse mit den Miniscaffoldinen, dass eine Immobilisierung keinen nachteiligen Effekt auf die hydrolytische Aktivität hatte.

Basierend auf diesen Gesichtspunkten und die bestehende Möglichkeit komplette Scaffoldin Proteine in den Komplex zu integrieren, könnte in naher Zukunft eine hoch effizientes, wiederverwendbares System entwickelt werden, welches mit den kommerziell erhältliches Cellulase Präparationen Schritt halten kann.

Literature

Adams, J. J., Webb, B. A., Spencer, H. L., Smith, S. P., (2005). Structural characterization of type II dockerin module from the cellulosome of *Clostridium thermocellum*: calcium-induced effects on conformation and target recognition. Biochemistry., 44(6):2173-82

Ahsan, M. M., Kimura, T., Karita, S., Sakka, K., Ohmiya, K., (1996). Cloning, DNA sequencing, and expression of the gene encoding *Clostridium thermocellum* cellulase CelJ, the largest catalytic component of the cellulosome. J Bacteriol., 178(19):5732-40

Ahsan, M., Matsumoto, M., Karita, S., Kimura, T., Sakka, K., Ohmiya, K., (1997). Purification and characterization of the family J catalytic domain derived from the *Clostridium thermocellum* endoglucanase CelJ. Biosci Biotechnol Biochem., 61(3):427-31

Alzari, P. M., Souchon, H., Dominguez, R. (1996). The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. Structure, 4(3):265-75.

Arai, T., Ohara, H., Karita, S., Kimura, T., Sakka, K., Ohmiya, K., (2001). Sequence of celQ and properties of celQ, a component of *the Clostridium thermocellum* cellulosome. Appl Microbiol Biotechnol., 57(5-6):660-6.

Arber, W., Linn, S., (1969). DNA modification and restriction. Annu Rev Biochem., 38:467-500

Bayer, E. A., Setter, E., Lamed, R., (1985). Organization and distribution of the cellulosome in *Clostridium thermocellum*. J. Bacteriol., 163:552-559

Bayer, E. A., Lamed R., (1986). Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. J Bacteriol., 167(3):828-36

Bayer, E. A., Shimon, L. J. W., Shoham, Y., Lamed, R, (1998). Cellulosomes - structure and ultrastructure. J. Struct. Biol., 124: 221–234

Barak, Y., Handelsman, T., Nakar, D., Mechaly, A., Lamed, R., Shoham, Y., Bayer, E. A., (2005). Matching fusion protein systems for affinity analysis of two interacting families of proteins: the cohesin-dockerin interaction. J. Mol. Recognit., 18(6):491-501

Béguin, P., Lemaire, M., (1996). The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. Crit. Rev. Biochem. Mol Biol., 31(3):201-36

Bender, J., Vatcharapijarn, Y., Jeffries, T. W., (1985). Characteristics and Adaptability of Some New Isolates of *Clostridium thermocellum*. Appl. Environ. Microbiol., 49(3):475-477.

Berger, E., Zhang, D., Zverlov, V. V., Schwarz, W. H., (2006). Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. FEMS Microbiol Lett., 268(2):194-201

Bhat, S., Goodenough, P. W., Bhat, M. K., Owen, E., (1994). Isolation of four major subunits from *Clostridium thermocellum* cellulosome and their synergism in the hydrolysis of crystalline cellulose. Int. J. Biol. Macromol., 16:335–342.

Bhat, M. K., (1998). *Clostridium thermocellum* cellulosome: Dissociation, isolation and characterisation of subunits and the potential biotechnological implications. Recent Res. Dev. Biotechnol. Bioeng., 1:59–84.

Blumer-Schuette, S. E., Kataeva, I., Westphe,ling, J., Adams, M. W., Kelly, R. M., (2008). Extremely thermophilic microorganisms for biomass conversion: status and prospects. Curr. Opin. Biotechnol., 19(3):210-217

Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P., Gilbert, H. J., (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem J., 331:775-81

Boraston, A. B., McLean, B. W., Guarna, M. M., Amandaron-Akow, E., Kilburn, D.G., (2001). A family 2a carbohydrate-binding module suitable as an affinity tag for proteins produced in *Pichia pastoris*. Protein Expr Purif., 21(3):417-23

Bradford, M. M., (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein dye binding. Anal. Biochem., 72:248-253

Carrard, G., Koivula, A., Söderlund, H., Béguin, P., (2000). Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. Proc Natl Acad Sci U S A, 97(19):10342-7

Carreira, L. H., Wiegel, J., Ljungdahl, L. G., (1983). Prodution of ethanol from biopolymers by anaerobic thermophilic and extreme thermophilic bacteria. I. Regulation of carbohydrate utilization in mutants of *Thermoanaerobacter ethanolicus*. Biotechnol. Bioeng. Symp. 13:183-191.

Caruso, F., (2000). Hollow capsule processing through colloidal templating and self-assembly. Chemistry, 6(3):413-9

Carvalho, A. L., Dias, F. M. V., Nagy, T., Prates, J. A. M., Proctor, M. R., Smith, N., Bayer, E. A., Davies, G. J., Ferreira, L. M. .A., Romao, M. J., Fontes, C. M. G. A., Gilbert, H. J, (2007). Evidence for a Dual Binding Mode of Dockerin Modules to Cohesins. Proc. Natl. Acad. Sci.USA 104: 3089

Caspi J., Barak Y., Haimovitz R., Irwin D., Lamed R., Wilson D. B., Bayer E. A., (2009). Effect of linker length and dockerin position on conversion of a *Thermobifida fusca* endoglucanase to the cellulosomal mode. Appl. Environ. Microbiol., 75(23):7335-42

Chauvaux, S., Béguin, P., Aubert, J. P., Bhat, K. M., Gow, L. A., Wood, T. M., Bairoch, A., (1990). Calcium-binding affinity and calcium-enhanced activity of *C lostridium thermocellum* endoglucanase D. Biochem J., 265(1):261-265

Chhabra, S. R., Kelly, R. M., (2002). Biochemical characterization of *Thermotoga maritima* endoglucanase Cel74 with and without a carbohydrate binding module (CBM). FEBS Lett., 531(2):375-80

Ciruela, A., Gilbert, H. J., Ali, B. R. S., Hazlewood, G. P., (1998). Synergistic interaction of the cellulosome integrating protein (CipA) from *Clostridium thermocellum* with a cellulosomal endoglucanase. FEBS Lett., 422:221–224.

Cox, P., Betts, R., Jones, C., Spall, S., Totterdell, I., (2000). Will carbon-cycle feedbacks accelerate global warming in the 21st century? Nature, 408: 184–187

Davies, G., Henrissat, B., (1995). Structures and mechanisms of glycosyl hydrolases. Structure, 15:853-859

Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Anthony, R. Jr., Warren, R. A. J., Kilburn, D. G., (1991). Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. Bio/Technol 9:1096– 1099

Farabaugh, P. J., (1978). Sequence of the *lacI* gene.Nature, 274(5673):765-9

Farrel, A. E., Plevin, R. J., Turner, B. T., Jones, A. D., O`Hare, M., Kammen, D. M., (2006). Ethanol can contribute to energy and environmental goals. Science, 311:506-508

Felix, C. R., Ljungdahl, L. G., (1993). The cellulosome: the exocellular organelle of *Clostridium*. Annu Rev Microbiol., 47:791-819

Fierobe, H. P., Pagès, S., Bélaïch, A., Champ, S., Lexa, D., Bélaïch, J. P., (1999). Cellulosome from *Clostridium cellulolyticum*: molecular study of the Dockerin/Cohesin interaction. Biochemistry, 38(39):12822-32

Fierobe, H. P., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J., Bayer, E. A., (2001). Design and Production of Active Cellulosome Chimeras – selective incorporation of dockerin-containing enzymes into defined functional complexes. J. Biol. Chem., 276(24):21257-21261

Fierobe, H. P., Bayer, E. A., Tardif, C., Czjzek, M., Mechaly, A., Bélaïch, A., Lamed, R., Shoham, Y., Bélaïch, J. P., (2002). Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. J. Biol. Chem., 277(51):49621-30

Fierobe, H. P., Mingardon, F., Mechaly, A., Bélaïch, A., Rincon, M. T., Pagès, S., Lamed, R., Tardif, C., Bélaïch, J. P., Bayer, E. A., (2005). Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. J. Biol. Chem., 280(16):16325-34

Freier, D., Mothershed, C. P., Wiegel, J., (1988). Characterization of *Clostridium thermocellum* JW20. Appl Environ Microbiol., 54(1):204-211.

Gold, N. D., Martin, V. J., (2007). Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. J Bacteriol., 189(19):6787-95

Goldemberg. J., (2007). Ethanol for a sustainable energy future. Science, 315:808-810

Gomez, L. D., Steele-King, C. G., McQueen-Mason, S. J., (2008). Sustainable liquid biofuels from biomass: the writing's on the walls. New Phytol., 178(3):473-85

Grépinet, O., Chebrou, M. C., Béguin, P., (1988). Purification of *Clostridium thermocellum* Xylanase Z expressed in *Escherichia coli* and identification of the corresponding product in the culture medium of *C. thermocellum*. J. Bacteriol., 170(10):4576-4581

Gilkes, N. R., Warren, R. A., Miller, R. C. J., Kilburn, D. G., (1988). Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem., 263:10401-10407

Haimovitz, R., Barak, Y., Morag, E., Voronov-Goldman, M., Shoham, Y., Lamed, R., Bayer, E. A., (2008). Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules. Proteomics, 8(5):968-79.

Halliwell, G., Griffin, M., (1973). The nature and mode of action of the cellulolytic component C1 of *Trichoderma koningii* on native cellulose. Biochem J., 135(4):587-94

Harhangi, H. R., Freelove, A. C., Ubhayasekera, W., van Dinther, M., Steenbakkers, P. J., Akhmanova, A., van der Drift, C., Jetten, M. S., Mowbray, S. L., Gilbert, H. J., Op den Camp, H. J., (2003). Cel6A, a major exoglucanase from the cellulosome of the anaerobic fungi *Piromyces sp.* E2 and *Piromyces equi*. Biochim Biophys Acta., 1628(1):30-9.

Hartmann, U., Mende, H. H., (1986). Hysteresis of Néel-line motion and effective width of 180 degrees Bloch walls in bulk iron. Phys Rev B Condens Matter, 33(7):4777-4781

Henrissat, B., Sulzenbacher, G., Bourne, Y., (2008). Glycosyltransferases, glycoside hydrolases: surprise, surprise! Curr. Opin. Struct. Biol., 18(5):527-33

Hermanson, G. T. (1995). Bioconjugate techniques. Academic Press

Howel, J. A., Stcuk, J. D., (1975). Kinetics of solka floc hydrolysis by *Trichoderma* cellulase. Biotechnol. Bioeng., 17:873-893

Inderwildi, O. R., King, D. A., Jenkins, S. J., (2009). Fischer-Tropsch synthesis of liquid fuels: learning lessons from homogeneous catalysis. Phys Chem Chem Phys., 11(47):11110-2

Johnson, E. A., Madia, A., Demain, A. L., (1981). Chemically defined minimal medium for growth of the anaerobic cellulolytic thermophile *Clostridium thermocellum*. Appl. Environ. Microbiol., 41 :1060-1062

Johnson, P. E., Tomme, P., Joshi, M. D., McIntosh, L. P., (1996). Interaction of soluble cellooligosaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC 2. NMR and ultraviolet absorption spectroscopy. Biochemistry., 35(44):13895-906

Joseleau, J. P., Comtat, J., Ruel, K., (1992). Chemical structure of xylans and their interaction in the plant cell walls. In: Xylans and Xylanases. Prog. Biotechnol., 7:1–15

Kadam, S. K., Demain, A. L., (1989). Addition of cloned beta-glucosidase enhances the degradation of crystalline cellulose by the *Clostridium thermocellum* cellulose complex. Biochem Biophys Res Commun., 161(2):706-11

Kataeva, I., Guglielmi, G., Béguin, P., (1997). Interaction between *Clostridium thermocellum* endoglucanase CelD and polypeptides derived from the cellulosome-integrating protein CipA: Stoichiometry and cellulolytic activity of the complexes. Biochem J., 326:617-624

Kataeva, I., Li, X. L., Chen, H., Choi, S. K., Ljungdahl, L. G., (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(17):5288-5295

Kato, Y., Nevins, D. J., (1985). Isolation and identification of *O*-(5-*O*-feruloyl-alpha-l-arabinofuranosyl)-(1-4)-d-xylopyranose as a component of *Zea* shoot cell-walls. Carbohydr. Res,, 137:139-150

Katz, M., Reese, E. T., (1968). Production of glucose by enzymatic hydrolysis of cellulose. Appl Microbiol., 16(2):419-20

Koeck, D., (2009). Untersuchung cellulolytischer Umweltisolate für den Einsatz zur Optimierung der Hydrolysestufe in Biogasanlagen. Master thesis, Department of Microbiology, TU München

Kouassi, G. K., Irudayaraj, J., McCarty, G., (2005). Examination of Cholesterol oxidase attachment onto magnetic nanoparticles. J. Nanobiotech, 3(1):1

Kruus, K., Andreacchi, A., Wang, W. K., Wu, J. H. D., (1995). Product inhibition of the recombinant CelS, an exoglucanase component of the *Clostridium thermocellum* cellulosome. Appl. Microbiol. Biotechnol., 44:399404

Lamed, R., Setter, E., Kenig, R., Bayer, E. A., (1983). The cellulosome: a discrete cell surface organelle of *Clostridium thermocellum* which inhibits separate antigenic, cellulose-binding and various cellulolytic activities. Biotechn. Bioeng. Symp., 13:163-181

Lamed, R., Kenig, R., Setter, E., Bayer, E. A., (1985). Major characteristics of the cellulolytic system of *Clostridium thermocellum* coincide with those of the purified cellulosome. Enzyme Microb. Techol., 7:37–41

Laemmli, U. K., (1970). Cleavage of structure proteins during the assembly of the head of bacteriophage T4. Nature, 227:680-685

Lee, H. J., Brown, R. M., (1997). A comparative structural characterization of two cellobiohydrolases from *Trichoderma reesei*: a high resolution electron microscopy study. Journal of Biotechnology, 57(16):127-136

Leschine, S. B., (1995). Cellulose degradation in anaerobic environments. Ann. Rev. Microbiol., 49: 399-426

Ljungdahl, L. G, Bryant, F., Carreira, L., Saiki, T., Wiegel, J., (1981). Some aspects of thermophilic and extreme thermophilic anaerobic microorganisms. Basic Life Sci., 18:397-419.

Luria, S. E., Adams, J. N., Ting, R. C., (1960). Transduction of lactose utilizing ability among strains of *E. coli* and *S. dysenterial* and the properties of the transducing phage particle. Virology, 12:348-390

Lynd, L. R., Laser, M. S., Bransby, D., Dale, B. E., Davison, B., Hamilton, R., Himmerl, M., Keller, M., McMillan, J. D., Sheehan, J., Wyman, C. E., (2008). How biotech can transform biofuels. Nat. Biotechnol., 26:169-172

Maugeri, F., Costa, A. C., Dechechi, E. C., Silva, F. L, Maciel, R. (1988). Simulated dynamics and control of an extractive alcoholic fermentation. Appl Biochem Biotechnol., 86:577-93

McBee, R. H., (1948). The culture and physiology of a thermophilic cellulose-fermenting bacterium. J. Bacteriol., 52 :653-663

McCarter, J. D., Withers, S. G., (1994). Mechanisms of enzymatic glycoside hydrolysis. Cur. Opp. Struct. Biol., 4 :885-892

Mingardon, F., Chanal, A., López-Contreras, A. M., Dray, C., Bayer, E. A., Fierobe, H. P., (2007). Incorporation of fungal cellulases in bacterial minicellulosomes yields viable, synergistically acting cellulolytic complexes. Appl. Environ. Microbiol., 73(12):3822-32

Miller, G. L., (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugars. Anal. Chem. 31:426-428

Morag, E., Bayer, E. A., Lamed, R., (1990). Relationship of cellulosomal and noncellulosomal xylanases of *Clostridium thermocellum* to cellulosedegrading enzymes. J. Bacteriol., 172:6098–6105

Morag, E., Halevy, I., Bayer, E. A., Lamed, R., (1991). Isolation and properties of a major cellobiohydrolase from the cellulosome of *Clostridium thermocellum*. J Bacteriol., 173(13):4155-62

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., (1986). Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol., 1:263-73

Murashima, K., Akihiko, K., Doi, R. H., (2003). Synergistic effects of cellulosomal xylanase and cellulases from *Clostridium cellulovorans* on plant cell wall degradation. J. Bacteriol., 185(5):1518-1524

Newcomb, M., Chen, C., and Wu, J. H, (2007). Induction of the celC operon of *Clostridium thermocellum* by laminaribiose. Proceedings of the National Academy of Sciences of the United States of America, 104:3747-3752.).

Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H., Yura, T., (1998). Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GroE and GroEL-GroES in assisting folding of an allergen of japanese cedar pollen, Cryj2, in *Escherichia coli*. Appl. Environ. Microbiol., 64:1694-1699

Nishihara, K., Kanemori, M., Yanagi, H., Yura, T., (2000). Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. Appl. Environ. Microbiol., 66:884-889

Ng, T. K., Weimer, T. K., Zeikus, J. G., (1977). Cellulolytic and physiological properties of *Clostridium thermocellum*. Arch Microbiol., 114(1):1-7

Ohmiya, K., Sakka, K., Karita, S., Kimura, T., (1997). Structure of cellulases and their applications. Biotechnol Genet Eng Rev., 14:365-414

Onda, M., Ariga, K., Kunitake, T., (1999). Activity and stability of glucose oxidase in molecular films assembled alternately with polyions. J Biosci Bioeng., 87(1):69-75

Pingoud, A., Jeltsch, A., (2001). Structure and function of type II restriction endonucleases. Nucleic Acids Res., 29(18):3705-27

Reese, E. T., Gilligan, W., Norkrans, B. (1952). Effect of cellobiose on the enzymatic hydrolysis of cellulose and its derivatives. Physiol. Plant., 5:379-390

Riedel, K., Bronnenmeier, L., (1998). Intramolecular synergism in a engineered endo-exo-1,4-β-glucanase fusion protein. Mol. Microbiol., 28(4):767-775

Riedel, K., Bronnenmeier, K., (1999). Active-site mutations which change the substrate specificity of the *Clostridium stercorarium* cellulase CelZ. Implications for synergism. Eur. J. Biochem. 262:218-223

Romaniec, M. P. M., Huskisson, N., Barker, P., Demain, A. L., (1992). Purification and properties of the *Clostridium thzermocellum bglB* gene product expressed in *Escherichia coli*. Enzyme Microb. Technol.,15:393-400

Sakon, J., Irwin, D., Wilson, D. B., Karplus, P. A., (1997). Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. Nat Struct Biol., (10):810-818

Schaeffer, F., Matuschek, M., Guglielmi, G., Miras, I., Alzari, P. M., Béguin, P., (2002). Duplicated dockerin subdomains of *Clostridium thermocellum* endoglucanase CelD bind to a cohesin domain of the scaffolding protein CipA with distinct thermodynamic parameters and a negative cooperativity. Biochemistry, 41(7):2106-14.

Schubot, F. D., Kataeva, I. A., Chang, J., Shah, A. K., Ljungdahl, L. G., Rose, J. P., Wang, B. C., (2004). Structural basis for the exocellulase activity of the cellobiohydrolase CbhA from *Clostridium thermocellum*. Biochemistry., 43(5):1163-70.

Schwarz, W. W., Gräbnitz, F., Staudenbauer, W. L., (1986). Properties of a *Clostridium thermocellum* endoglucanase produced in *Escherichia coli*. Appl. Env. Microbiology, 51(6):1293-1299

Schwarz, W. H., (2003). Das Cellulosom – eine Nano-Maschine zum Abbau von Cellulose. NR., 56: 121-128

Schwarz, W. H., (2004). Cellulose – Struktur ohne Ende. NR., 57: 443-445

Schwarz, W. H., (2001). The cellulosome and cellulose degradation by anaerobic bacteria. Appl. Microb. Biot., 56(5-6):634-49

Schwarz, W. H, Antoni, D., Zverlov, V V., (2007). Biofuels from microbes. Appl. Microbiol. Biotechnol., 77(1):23-35

Shieh, D. B., Su, C. H., Chang, F. Y., Wu, Y. N., Su, W. C., Hwu, J. R., Chen, J. H., Yeh, C. S., (2006). Aqueous nickel-nitrilotriacetate modified Fe_3O_4–NH^{+3} nanoparticles for protein purification and cell targeting. Nanotechnology, 17:4174-4182

Shuman S., (1991). Recombination mediated by vaccinia virus DNA topoisomerase I in *Escherichia coli* is sequence specific. Proc Natl Acad Sci USA., 88(22):10104-8

Shuman, S., (1994). Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. J Biol Chem., 269(51):32678-84

Sommerville, C., (2008). Development of cellulosic biofuels. U.S. Department of Agriculture.

Staros, J. V., Wright, R. W., Swingle, D. M., (1986). Enhancement by *N*-hydroxysulfosuccinimide of water-soluble carbodiimide-mediated coupling reactions. Anal. Biochem., 156:220-222

Teeri, T. T., Srisodsuk, M., Lehtiö, J., Linder, M., Margolles-Clark, E., Reinikainen, T., (1997). *Trichoderma reesei* cellobiohydrolase I with an endoglucanase cellulose-binding domain: action on bacterial microcrystalline cellulose. J Biotechnol., 57(1-3):49-57.

Tokatlidis, K., Salamitou, S., Béguin, P., Dhurjati, P., Aubert, J. P., (1991). Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. FEBS Lett., 291(2):185-8

Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., Claeyssens, M., (1988). Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of

domain function in two cellobiohydrolases by limited proteolysis. Eur. J. Biochem., 170:575-581

Tomme, P., Warren, R. A. J., Gilkes, N. R., (1995). Cellulose hydrolysis by bacteria and fungi. Adv. In Microbial Physiol., 37: 1-81

Tomme, P., Gilkes, N. R., Guarna, M. M., Haynes, C. A., Hasenwinkle, D., Jervis, E., Johnson, P., McIntosh, L., Warren, R. A., Kilburn, D. G., (1996). Cellulose-binding domains. Versatile affinity tags for inexpensive large-scale purification, concentration, and immobilization of fusion proteins. Ann N Y Acad Sci., 799:418-24

Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A. L., Sturch, K., Gilkes, N. R., Haynes. C. A., Warren, R. A., Kilburn, D. G., (1998). Characterization and affinity applications of cellulose-binding domains. J Chromatogr B Biomed Sci Appl., 715(1):283-96

van Ketel, W. G., Bruynzeel, D. P., (1984). Chelating effect of EDTA on nickel. Contact Dermatitis., 11(5):311-4

van Tilbeurgh, H., Pettersson, G., Bhikabhai, R., De Boeck, H., Claeyssens, M., (1985). Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Reaction specificity and thermodynamics of interactions of small substrates and ligands with the 1,4-beta-glucan cellobiohydrolase II. Eur J Biochem., 148(2):329-34

Viljoen, J. A., Fred, E. B., Peterson, W. H., (1926). *Clostridium thermocellum* J. of Agric. Sc., 16, 1

Vocadlo, D. J., Davies, G. J., Laine, R., Withers, S. G., (2001). Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. Nature 412, 835.

Wang, W. K., Kruus, K., Wu, J. H., (1993). Cloning and DNA sequence of the gene coding for *Clostridium thermocellum* cellulase Ss (CelS), a major cellulosome component. J Bacteriol., 175(5):1293-302

Wang, W. K., Kruus, K., Wu, J. H., (1994). Cloning and expression of the *Clostridium thermocellum* celS gene in *Escherichia coli*. Appl Microbiol Biotechnol., 42(2-3):346-52

Wang, P., Dai, S., Waezsada, S. D., Tsao, A. Y., Davison, B. H., (2001). Enzyme stabilization by covalent binding in nanoporous sol-gel glass for nonaqueous biocatalysis. Biotechnol Bioeng., 74(3):249-55

Wiegel, J., Mothershed, C. P., Puls, J., (1985). Differences in xylan degradation by various noncellulolytic thermophilic anaerobes and *Clostridium thermocellum*. Appl Environ Microbiol., 49(3):656-659.

Wiegel, J., Ljungdahl, L. G.,Demain, L. A., (1985). The importance of thermophilic bacteria in biotechnology. Critical Reviews in Biotechnology 3(1):39-108

Wong, K., Tan, L., Saddler, J. N., (1988). Multiplicity of b-1,4-xylanase in microorganisms: functions and applications. Microbiol Rev, 52:305–317

Wood, T. M., McCrae, S. I., (1978). Bioconversion of cellulosic substances into energy chemicals and microbial protein. Indian Institute of Technology, New Dehli. P. 111-142

Wu, J. H. D., Orme-Johnson, W. H., Demain, A. L., (1988). Two components of an extracellular protein aggregate of *Clostridium thermocellum* together degrade crystalline cellulose. Biochemistry, 27:1703–1709

Xiao, Z., Zhang, X., Gregg, D. J., Saddler, J. N., (2004). Effects of sugar inhibition on cellulases and beta-glucosidase during enzymatic hydrolysis of softwood substrates. Appl Biochem Biotechnol., 113-116:1115-26

Xu, Q., Gao, W., Ding, S. Y., Kenig, R., Shoham, Y., Bayer, E. A., Lamed, R., (2003). The cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell surface anchoring protein. J Bacteriol., 185(15):4548-57

Xu, Q., Singh, A., Himmel, M. E., (2009). Perspective and new directions for the production of bioethanol using consolidated bioprocessing of lignocellulose. Curr. Opin. Biotechnol., 20:364-371

Young, M., Minton, N. P., Staudenbauer, W. L., (1989). Recent advances in the genetics of the clostridia. FEMS Microbiol. Rev., 63:301-326

Zhang, Y., Lynd, L. R., (2003). Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: development of an enzyme-linked immunosorbent assay-based method with application to *Clostridium thermocellum* batch cultures. Anal. Chem. 75:219–227

Zverlov, V. V., Volkov, I. Y., Velikodvorskaya, T. V., Schwarz, W. H., (1997). *Thermotoga neapolitana* bglB gene, upstream of lamA, encodes a highly thermostable beta-glucosidase that is a laminaribiase. Microbiology, 143:3537-42.

Zverlov, V. V., Mahr, S., Riedel, K., Bronnenmeier K., (1998). Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile *Anaerocellum thermophilum* with separate glycosyl hydrolase family 9 and 48 catalytic domains. Microbiology, 144:457-65

Zverlov, V. V., Schantz, N., Schwarz, W. H., (2005). A major new component in the cellulosome of *Clostridium thermocellum* is a processive

endo-beta-1,4-glucanase producing cellotetraose. FEMS Microbiol Lett., 249(2):353-8.

Zverlov, V. V., Schantz, N., Schmitt-Kopplin, P., Schwarz, W. H., (2005). Two new major subunits in the cellulosome of *Clostridium thermocellum*: xyloglucanase Xgh74A and endoxylanase Xyn10D. Microbiology, 151:3395-401

Zverlov, V. V., Kellermann, J., Schwarz, W. H., (2005). Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. Proteomics, 5(14):3646-53

Zverlov, V. V, Berezina, O., Velikodvorskaya, G. A, Schwarz, W. H., (2006). Bacterial acetone and butanol production by industrial fermentation in the Soviet Union: use of hydrolyzed agricultural waste for biorefinery. Appl. Microbiol. Biotechnol., 71(5):587-97

Zverlov, V. V., Schwarz, W. H., (2008). Bacterial cellulose hydrolysis in anaerobic environmental subsystems--*Clostridium thermocellum* and *Clostridium stercorarium*, thermophilic plant-fiber degraders. Ann N Y Acad Sci., 1125:298-307

Zverlov, V. V., Klupp, M., Krauss, J., Schwarz, W. H., (2008). Mutations in the scaffoldin gene, cipA, of *Clostridium thermocellum* with impaired cellulosome formation and cellulose hydrolysis: insertions of a new transposable element, IS1447, and implications for cellulase synergism on crystalline cellulose. J. Bacteriol., 190(12):4321-7

Appendix

DNA sequences of cloned genes

Red marked: start codon; squared: putative promotor; underlined: Shine-Dalgarno sequence, dotted: stop codon; grey marked: cohesin type I module; blue marked: CBM, rose marked: dockerin type II module;

Scaffoldin protein CipA

```
  1  ATA TTG TTT ATT AAT ATT AAC ACT AAT TTT TGT TAT TGT CTT GCT TGG TTT
       I   L   F   I   N   I   N   T   N   F   C   Y   C   L   A   W   F

 52  GTA TAT AAG GTA TTT GAT TTC AAA TGC CTT TAG AGA TCT TTT TAT AAG ATC
       V   Y   K   V   F   D   F   K   C   L   -   R   S   F   Y   K   I

103  ATA TAA AAT ATT AAT TTT TGG GAG GAA TGG TAG ATG AGA AAA GTC ATC AGT
       I   -   N   I   N   F   W   E   E   W   -   M   R   K   V   I   S

154  ATG CTC TTA GTT GTG GCT ATG CTG ACG ACG ATT TTT GCG GCG ATG ATA CCG
       M   L   L   V   V   A   M   L   T   T   I   F   A   A   M   I   P

205  CAG ACA GTA TCG GCG GCC ACA ATG ACA GTC GAG ATC GGC AAA GTT ACA GCA
       Q   T   V   S   A   A   T   M   T   V   E   I   G   K   V   T   A

256  GCC GTT GGA TCA AAA GTA GAA ATA CCT ATA ACC CTG AAA GGA GTG CCA TCC
       A   V   G   S   K   V   E   I   P   I   T   L   K   G   V   P   S

307  AAA GGA ATG GCC AAT TGC GAC TTC GTA TTG GGT TAT GAT CCA AAT GTG CTG
       K   G   M   A   N   C   D   F   V   L   G   Y   D   P   N   V   L

358  GAA GTA ACA GAA GTA AAA CCA GGA AGC ATA ATA AAA GAT CCG GAT CCT AGC
       E   V   T   E   V   K   P   G   S   I   I   K   D   P   D   P   S

409  AAG AGC TTT GAT AGC GCA ATA TAT CCG GAT CGA AAG ATG ATT GTA TTT CTG
       K   S   F   D   S   A   I   Y   P   D   R   K   M   I   V   F   L

460  TTT GCA GAA GAC AGT GGA AGA GGA ACG TAT GCA ATA ACT CAG GAT GGA GTA
       F   A   E   D   S   G   R   G   T   Y   A   I   T   Q   D   G   V

511  TTT GCA ACA ATT GTA GCC ACT GTC AAA TCA GCT GCA GCG GCA CCG ATT ACT
       F   A   T   I   V   A   T   V   K   S   A   A   A   A   P   I   T

562  TTG CTT GAA GTA GGT GCA TTT GCG GAC AAC GAT TTA GTA GAA ATA AGC ACA
       L   L   E   V   G   A   F   A   D   N   D   L   V   E   I   S   T

613  ACT TTT GTC GCG GGC GGA GTA AAT CTT GGT AGT TCC GTA CCG ACA ACA CAG
       T   F   V   A   G   G   V   N   L   G   S   S   V   P   T   T   Q

664  CCA AAT GTT CCG TCA GAC GGT GTG GTA GTA GAA ATT GGC AAA GTT ACG GGA
       P   N   V   P   S   D   G   V   V   V   E   I   G   K   V   T   G


715  TCT GTT GGA ACT ACA GTT GAA ATA CCT GTA TAT TTC AGA GGA GTT CCA TCC
       S   V   G   T   T   V   E   I   P   V   Y   F   R   G   V   P   S
```

```
 766  AAA GGA ATA GCA AAC TGC GAC TTT GTG TTC AGA TAT GAT CCG AAT GTA TTG
        K   G   I   A   N   C   D   F   V   F   R   Y   D   P   N   V   L

 817  GAA ATT ATA GGG ATA GAT CCC GGA GAC ATA ATA GTT GAC CCG AAT CCT ACC
        E   I   I   G   I   D   P   G   D   I   I   V   D   P   N   P   T

 868  AAG AGC TTT GAT ACT GCA ATA TAT CCT GAC AGA AAG ATA ATA GTA TTC CTG
        K   S   F   D   T   A   I   Y   P   D   R   K   I   I   V   F   L

 919  TTT GCG GAA GAC AGC GGA ACA GGA GCG TAT GCA ATA ACT AAA GAC GGA GTA
        F   A   E   D   S   G   T   G   A   Y   A   I   T   K   D   G   V

 970  TTT GCA AAA ATA AGA GCA ACT GTA AAA TCA AGT GCT CCG GGC TAT ATT ACT
        F   A   K   I   R   A   T   V   K   S   S   A   P   G   Y   I   T

1021  TTC GAC GAA GTA GGT GGA TTT GCA GAT AAT GAC CTG GTA GAA CAG AAG GTA
        F   D   E   V   G   G   F   A   D   N   D   L   V   E   Q   K   V

1072  TCA TTT ATA GAC GGT GGT GTT AAC GTT GGC AAT GCA ACA CCG ACC AAG GGA
        S   F   I   D   G   G   V   N   V   G   N   A   T   P   T   K   G

1123  GCA ACA CCA ACA AAT ACA GCT ACG CCG ACA AAA TCA GCT ACG GCT ACG CCC
        A   T   P   T   N   T   A   T   P   T   K   S   A   T   A   T   P

1174  ACC AGG CCA TCG GTA CCG ACA AAC ACA CCG ACA AAC ACA CCG GCA AAT ACA
        T   R   P   S   V   P   T   N   T   P   T   N   T   P   A   N   T

1225  CCG GTA TCA GGC AAT TTG AAG GTT GAA TTC TAC AAC AGC AAT CCT TCA GAT
        P   V   S   G   N   L   K   V   E   F   Y   N   S   N   P   S   D

1276  ACT ACT AAC TCA ATC AAT CCT CAG TTC AAG GTT ACT AAT ACC GGA AGC AGT
        T   T   N   S   I   N   P   Q   F   K   V   T   N   T   G   S   S

1327  GCA ATT GAT TTG TCC AAA CTC ACA TTG AGA TAT TAT TAT ACA GTA GAC GGA
        A   I   D   L   S   K   L   T   L   R   Y   Y   Y   T   V   D   G

1378  CAG AAA GAT CAG ACC TTC TGG TGT GAC CAT GCT GCA ATA ATC GGC AGT AAC
        Q   K   D   Q   T   F   W   C   D   H   A   A   I   I   G   S   N

1429  GGC AGC TAC AAC GGA ATT ACT TCA AAT GTA AAA GGA ACA TTT GTA AAA ATG
        G   S   Y   N   G   I   T   S   N   V   K   G   T   F   V   K   M

1480  AGT TCC TCA ACA AAT AAC GCA GAC ACC TAC CTT GAA ATA AGC TTT ACA GGC
        S   S   S   T   N   N   A   D   T   Y   L   E   I   S   F   T   G

1531  GGA ACT CTT GAA CCG GGT GCA CAT GTT CAG ATA CAA GGT AGA TTT GCA AAG
        G   T   L   E   P   G   A   H   V   Q   I   Q   G   R   F   A   K

1582  AAT GAC TGG AGT AAC TAT ACA CAG TCA AAT GAC TAC TCA TTC AAG TCT GCT
        N   D   W   S   N   Y   T   Q   S   N   D   Y   S   F   K   S   A

1633  TCA CAG TTT GTT GAA TGG GAT CAG GTA ACA GCA TAC TTG AAC GGT GTT CTT
        S   Q   F   V   E   W   D   Q   V   T   A   Y   L   N   G   V   L

1684  GTA TGG GGT AAA GAA CCC GGT GGC AGT GTA GTA CCA TCA ACA CAG CCT GTA
        V   W   G   K   E   P   G   G   S   V   V   P   S   T   Q   P   V

1735  ACA ACA CCA CCT GCA ACA ACA AAA CCA CCT GCA ACA ACA AAA CCA CCT GCA
        T   T   P   P   A   T   T   K   P   P   A   T   T   K   P   P   A

1786  ACA ACA ATA CCG CCG TCA GAT GAT CCG AAT GCA ATA AAG ATT AAG GTG GAC
        T   T   I   P   P   S   D   D   P   N   A   I   K   I   K   V   D
```

```
1837  ACA GTA AAT GCA AAA CCG GGA GAC ACA GTA AAT ATA CCT GTA AGA TTC AGT
        T   V   N   A   K   P   G   D   T   V   N   I   P   V   R   F   S

1888  GGT ATA CCA TCC AAG GGA ATA GCA AAC TGT GAC TTT GTA TAC AGC TAT GAC
        G   I   P   S   K   G   I   A   N   C   D   F   V   Y   S   Y   D

1939  CCG AAT GTA CTT GAG ATA ATA GAG ATA AAA CCG GGA GAA TTG ATA GTT GAC
        P   N   V   L   E   I   I   E   I   K   P   G   E   L   I   V   D

1990  CCG AAT CCT GAC AAG AGC TTT GAT ACT GCA GTA TAT CCT GAC AGA AAG ATA
        P   N   P   D   K   S   F   D   T   A   V   Y   P   D   R   K   I

2041  ATA GTA TTC CTG TTT GCA GAA GAC AGC GGA ACA GGA GCG TAT GCA ATA ACT
        I   V   F   L   F   A   E   D   S   G   T   G   A   Y   A   I   T

2092  AAA GAC GGA GTA TTT GCT ACG ATA GTA GCG AAA GTA AAA TCC GGA GCA CCT
        K   D   G   V   F   A   T   I   V   A   K   V   K   S   G   A   P

2143  AAC GGA CTC AGT GTA ATC AAA TTT GTA GAA GTA GGC GGA TTT GCG AAC AAT
        N   G   L   S   V   I   K   F   V   E   V   G   G   F   A   N   N

2194  GAC CTT GTA GAA CAG AGG ACA CAG TTC TTT GAC GGT GGA GTA AAT GTT GGA
        D   L   V   E   Q   R   T   Q   F   F   D   G   G   V   N   V   G

2245  GAT ACA ACA GTA CCT ACA ACA CCT ACA ACA CCT GTA ACA ACA CCG ACA GAT
        D   T   T   V   P   T   T   P   T   T   P   V   T   T   P   T   D

2296  GAT TCG AAT GCA GTA AGG ATT AAG GTG GAC ACA GTA AAT GCA AAA CCG GGA
        D   S   N   A   V   R   I   K   V   D   T   V   N   A   K   P   G

2347  GAC ACA GTA AGA ATA CCT GTA AGA TTC AGC GGT ATA CCA TCC AAG GGA ATA
        D   T   V   R   I   P   V   R   F   S   G   I   P   S   K   G   I

2398  GCA AAC TGT GAC TTT GTA TAC AGC TAT GAC CCG AAT GTA CTT GAG ATA ATA
        A   N   C   D   F   V   Y   S   Y   D   P   N   V   L   E   I   I

2449  GAG ATA GAA CCG GGA GAC ATA ATA GTT GAC CCG AAT CCT GAC AAG AGC TTT
        E   I   E   P   G   D   I   I   V   D   P   N   P   D   K   S   F

2500  GAT ACT GCA GTA TAT CCT GAC AGA AAG ATA ATA GTA TTC CTG TTT GCG GAA
        D   T   A   V   Y   P   D   R   K   I   I   V   F   L   F   A   E

2551  GAC AGC GGA ACA GGA GCG TAT GCA ATA ACT AAA GAC GGA GTA TTT GCT ACG
        D   S   G   T   G   A   Y   A   I   T   K   D   G   V   F   A   T

2602  ATA GTA GCG AAA GTA AAA TCC GGA GCA CCT AAC GGA CTC AGT GTA ATC AAA
        I   V   A   K   V   K   S   G   A   P   N   G   L   S   V   I   K

2653  TTT GTA GAA GTA GGC GGA TTT GCG AAC AAT GAC CTT GTA GAA CAG AAG ACA
        F   V   E   V   G   G   F   A   N   N   D   L   V   E   Q   K   T

2704  CAG TTC TTT GAC GGT GGA GTA AAT GTT GGA GAT ACA ACA GAA CCT GCA ACA
        Q   F   F   D   G   G   V   N   V   G   D   T   T   E   P   A   T

2755  CCT ACA ACA CCT GTA ACA ACA CCG ACA ACA ACA GAT GAT CTG GAT GCA GTA
        P   T   T   P   V   T   T   P   T   T   T   D   D   L   D   A   V

2806  AGG ATT AAA GTG GAC ACA GTA AAT GCA AAA CCG GGA GAC ACA GTA AGA ATA
        R   I   K   V   D   T   V   N   A   K   P   G   D   T   V   R   I

2857  CCT GTA AGA TTC AGC GGT ATA CCA TCC AAG GGA ATA GCA AAC TGT GAC TTT
        P   V   R   F   S   G   I   P   S   K   G   I   A   N   C   D   F
```

```
2908  GTA TAC AGC TAT GAC CCG AAT GTA CTT GAG ATA ATA GAG ATA GAA CCG GGA
        V   Y   S   Y   D   P   N   V   L   E   I   I   E   I   E   P   G

2959  GAC ATA ATA GTT GAC CCG AAT CCT GAC AAG AGC TTT GAT ACT GCA GTA TAT
        D   I   I   V   D   P   N   P   D   K   S   F   D   T   A   V   Y

3010  CCT GAC AGA AAG ATA ATA GTA TTC CTG TTT GCG GAA GAC AGC GGA ACA GGA
        P   D   R   K   I   I   V   F   L   F   A   E   D   S   G   T   G

3061  GCG TAT GCA ATA ACT AAA GAC GGA GTA TTT GCT ACG ATA GTA GCG AAA GTA
        A   Y   A   I   T   K   D   G   V   F   A   T   I   V   A   K   V

3112  AAA TCC GGA GCA CCT AAC GGA CTC AGT GTA ATC AAA TTT GTA GAA GTA GGC
        K   S   G   A   P   N   G   L   S   V   I   K   F   V   E   V   G

3163  GGA TTT GCG AAC AAT GAC CTT GTA GAA CAG AAG ACA CAG TTC TTT GAC GGT
        G   F   A   N   N   D   L   V   E   Q   K   T   Q   F   F   D   G

3214  GGA GTA AAT GTT GGA GAT ACA ACA GAA CCT GCA ACA CCT ACA ACA CCT GTA
        G   V   N   V   G   D   T   T   E   P   A   T   P   T   T   P   V

3265  ACA ACA CCG ACA ACA ACA GAT GAT CTG GAT GCA GTA AGG ATT AAA GTG GAC
        T   T   P   T   T   T   D   D   L   D   A   V   R   I   K   V   D

3316  ACA GTA AAT GCA AAA CCG GGA GAC ACA GTA AGA ATA CCT GTA AGA TTC AGC
        T   V   N   A   K   P   G   D   T   V   R   I   P   V   R   F   S

3367  GGT ATA CCA TCC AAG GGA ATA GCA AAC TGT GAC TTT GTA TAC AGC TAT GAC
        G   I   P   S   K   G   I   A   N   C   D   F   V   Y   S   Y   D

3418  CCG AAT GTA CTT GAG ATA ATA GAG ATA GAA CCG GGA GAC ATA ATA GTT GAC
        P   N   V   L   E   I   I   E   I   E   P   G   D   I   I   V   D

3469  CCG AAT CCT GAC AAG AGC TTT GAT ACT GCA GTA TAT CCT GAC AGA AAG ATA
        P   N   P   D   K   S   F   D   T   A   V   Y   P   D   R   K   I

3520  ATA GTA TTC CTG TTT GCA GAA GAC AGC GGA ACA GGA GCG TAT GCA ATA ACT
        I   V   F   L   F   A   E   D   S   G   T   G   A   Y   A   I   T

3571  AAA GAC GGA GTA TTT GCT ACG ATA GTA GCG AAA GTA AAA GAA GGA GCA CCT
        K   D   G   V   F   A   T   I   V   A   K   V   K   E   G   A   P

3622  AAC GGA CTC AGT GTA ATC AAA TTT GTA GAA GTA GGC GGA TTT GCG AAC AAT
        N   G   L   S   V   I   K   F   V   E   V   G   G   F   A   N   N

3673  GAC CTT GTA GAA CAG AAG ACA CAG TTC TTT GAC GGT GGA GTA AAT GTT GGA
        D   L   V   E   Q   K   T   Q   F   F   D   G   G   V   N   V   G

3724  GAT ACA ACA GAA CCT GCA ACA CCT ACA ACA CCT GTA ACA ACA CCG ACA ACA
        D   T   T   E   P   A   T   P   T   T   P   V   T   T   P   T   T

3775  ACA GAT GAT CTG GAT GCA GTA AGG ATT AAA GTG GAC ACA GTA AAT GCA AAA
        T   D   D   L   D   A   V   R   I   K   V   D   T   V   N   A   K

3826  CCG GGA GAC ACA GTA AGA ATA CCT GTA AGA TTC AGC GGT ATA CCA TCC AAG
        P   G   D   T   V   R   I   P   V   R   F   S   G   I   P   S   K

3877  GGA ATA GCA AAC TGT GAC TTT GTA TAC AGC TAT GAC CCG AAT GTA CTT GAG
        G   I   A   N   C   D   F   V   Y   S   Y   D   P   N   V   L   E

3928  ATA ATA GAG ATA GAA CCG GGA GAA TTG ATA GTT GAC CCG AAT CCT ACC AAG
        I   I   E   I   E   P   G   E   L   I   V   D   P   N   P   T   K
```

```
3979  AGC TTT GAT ACT GCA GTA TAT CCT GAC AGA AAG ATG ATA GTA TTC CTG TTT
        S   F   D   T   A   V   Y   P   D   R   K   M   I   V   F   L   F

4030  GCG GAA GAC AGC GGA ACA GGA GCG TAT GCA ATA ACT GAA GAT GGA GTA TTT
        A   E   D   S   G   T   G   A   Y   A   I   T   E   D   G   V   F

4081  GCT ACG ATA GTA GCG AAA GTA AAA TCC GGA GCA CCT AAC GGA CTC AGT GTA
        A   T   I   V   A   K   V   K   S   G   A   P   N   G   L   S   V

4132  ATC AAA TTT GTA GAA GTA GGC GGA TTT GCG AAC AAT GAC CTT GTA GAA CAG
        I   K   F   V   E   V   G   G   F   A   N   N   D   L   V   E   Q

4183  AAG ACA CAG TTC TTT GAC GGT GGA GTA AAT GTT GGA GAT ACA ACA GAA CCT
        K   T   Q   F   F   D   G   G   V   N   V   G   D   T   T   E   P

4234  GCA ACA CCT ACA ACA CCT GTA ACA ACA CCG ACA ACA ACA GAT GAT CTG GAT
        A   T   P   T   T   P   V   T   T   P   T   T   T   D   D   L   D

4285  GCA GTA AGG ATT AAA GTG GAC ACA GTA AAT GCA AAA CCG GGA GAC ACA GTA
        A   V   R   I   K   V   D   T   V   N   A   K   P   G   D   T   V

4336  AGA ATA CCT GTA AGA TTC AGC GGT ATA CCA TCC AAG GGA ATA GCA AAC TGT
        R   I   P   V   R   F   S   G   I   P   S   K   G   I   A   N   C

4387  GAC TTT GTA TAC AGC TAT GAC CCG AAT GTA CTT GAG ATA ATA GAG ATA GAA
        D   F   V   Y   S   Y   D   P   N   V   L   E   I   I   E   I   E

4438  CCG GGA GAC ATA ATA GTT GAC CCG AAT CCT GAC AAG AGC TTT GAT ACT GCA
        P   G   D   I   I   V   D   P   N   P   D   K   S   F   D   T   A

4489  GTA TAT CCT GAC AGA AAG ATA ATA GTA TTC CTG TTT GCA GAA GAC AGC GGA
        V   Y   P   D   R   K   I   I   V   F   L   F   A   E   D   S   G

4540  ACG GGA GCG TAT GCA ATA ACT AAA GAC GGA GTA TTT GCT ACG ATA GTA GCG
        T   G   A   Y   A   I   T   K   D   G   V   F   A   T   I   V   A

4591  AAA GTA AAA GAA GGA GCA CCT AAC GGA CTC AGT GTA ATC AAA TTT GTA GAA
        K   V   K   E   G   A   P   N   G   L   S   V   I   K   F   V   E

4642  GTA GGC GGA TTT GCG AAC AAT GAC CTT GTA GAA CAG AAG ACA CAG TTC TTT
        V   G   G   F   A   N   N   D   L   V   E   Q   K   T   Q   F   F

4693  GAC GGT GGA GTA AAT GTT GGA GAT ACA ACA GTA CCT ACA ACA TCG CCG ACA
        D   G   G   V   N   V   G   D   T   T   V   P   T   T   S   P   T

4744  ACA ACA CCG CCA GAG CCG ACG ATA ACT CCG AAC AAG TTG ACA CTT AAG ATA
        T   T   P   P   E   P   T   I   T   P   N   K   L   T   L   K   I

4795  GGC AGA GCA GAA GGA AGA CCT GGA GAC ACG GTG GAA ATA CCG GTT AAC TTG
        G   R   A   E   G   R   P   G   D   T   V   E   I   P   V   N   L

4846  TAT GGA GTA CCT CAA AAA GGA ATA GCA AGC GGT GAC TTC GTA GTA AGC TAT
        Y   G   V   P   Q   K   G   I   A   S   G   D   F   V   V   S   Y

4897  GAC CCG AAT GTA CTT GAG ATA ATA GAG ATA GAA CCG GGA GAA TTG ATA GTT
        D   P   N   V   L   E   I   I   E   I   E   P   G   E   L   I   V

4948  GAC CCG AAT CCT ACC AAG AGC TTT GAT ACT GCA GTA TAT CCT GAC AGA AAG
        D   P   N   P   T   K   S   F   D   T   A   V   Y   P   D   R   K

4999  ATG ATA GTA TTC CTG TTT GCG GAA GAC AGC GGA ACA GGA GCG TAT GCA ATA
        M   I   V   F   L   F   A   E   D   S   G   T   G   A   Y   A   I
```

```
5050  ACT GAA GAT GGA GTA TTT GCT ACG ATA GTA GCG AAA GTA AAA GAA GGA GCA
        T   E   D   G   V   F   A   T   I   V   A   K   V   K   E   G   A

5101  CCT GAA GGA TTC AGT GCA ATA GAA ATT TCT GAG TTT GGT GCA TTT GCA GAT
        P   E   G   F   S   A   I   E   I   S   E   F   G   A   F   A   D

5152  AAT GAT CTG GTA GAA GTG GAA ACT GAC CTT ATC AAT GGT GGA GTA CTT GTA
        N   D   L   V   E   V   E   T   D   L   I   N   G   G   V   L   V

5203  ACT AAT AAA CCT GTA ATA GAA GGA TAT AAA GTA TCC GGA TAC ATT TTG CCA
        T   N   K   P   V   I   E   G   Y   K   V   S   G   Y   I   L   P

5254  GAC TTC TCC TTC GAC GCT ACT GTT GCA CCA CTT GTA AAG GCC GGA TTC AAA
        D   F   S   F   D   A   T   V   A   P   L   V   K   A   G   F   K

5305  GTT GAA ATA GTA GGA ACA GAA TTG TAT GCA GTA ACA GAT GCA AAC GGA TAC
        V   E   I   V   G   T   E   L   Y   A   V   T   D   A   N   G   Y

5356  TTT GAA ATA ACC GGA GTA CCT GCA AAT GCA AGC GGA TAT ACA TTG AAG ATT
        F   E   I   T   G   V   P   A   N   A   S   G   Y   T   L   K   I

5407  TCA AGA GCA ACT TAC TTG GAC AGA GTA ATT GCA AAT GTT GTA GTA ACG GGA
        S   R   A   T   Y   L   D   R   V   I   A   N   V   V   V   T   G

5458  GAT ACT TCA GTT TCA ACT TCA CAG GCT CCA ATA ATG ATG TGG GTA GGA GAC
        D   T   S   V   S   T   S   Q   A   P   I   M   M   W   V   G   D

5509  ATA GTG AAA GAC AAT TCT ATC AAC CTG TTG GAC GTT GCA GAA GTT ATC CGT
        I   V   K   D   N   S   I   N   L   L   D   V   A   E   V   I   R

5560  TGC TTC AAC GCT ACT AAA GGA AGC GCA AAC TAC GTA GAA GAA CTT GAC ATT
        C   F   N   A   T   K   G   S   A   N   Y   V   E   E   L   D   I

5611  AAT AGA AAC GGC GCA ATT AAC ATG CAA GAC ATA ATG ATT GTT CAT AAG CAC
        N   R   N   G   A   I   N   M   Q   D   I   M   I   V   H   K   H
                                                    •
5662  TTT GGA GCT ACA TCA AGT GAT TAC GAC GCA CAG TAA ATA TTA AAA TTG GGA
        F   G   A   T   S   S   D   Y   D   A   Q   -   I   L   K   L   G

5713  GGA AGG ATA CCC CCC GGT ATC CTT CCT CTC AAA AAT ATT CTT TTT TTA TAT
        G   R   I   P   P   G   I   L   P   L   K   N   I   L   F   L   Y

5764  TTG AAA AGC AGA AAG AGA GAA ACA GAT TAA AAA TTA GAG CTA TAT GTG CTA
        L   K   S   R   K   R   E   T   D   -   K   L   E   L   Y   V   L

5815  TAC ATG AGC TGT TGA AGG GGG GAA TTT TTT CTT CAT GAA ACG AAA AAA TAA
        Y   M   S   C   -   R   G   E   F   F   L   H   E   T   K   K   -
```

Cellobiohydrolase A (GH9)

```
  1  TTG AAA AAT ACT TGA CAT TAC TGG TTG ACA TTT TAT ATA TTT AAA GTA TAA
       L   K   N   T   -   H   Y   W   L   T   F   Y   I   F   K   V   -

 52  GGA ACA TAA TTT GCA CCG ATC CTG TTC TAA ATT CTG AAA TTA ACA AGT ATT
       G   T   -   F   A   P   I   L   F   -   I   L   K   L   T   S   I

103  TGG GGT TGT ATT ATA TAT GGG GTT GTA TCA TTG GGG TAT TAT AAT CAA TGC
       W   G   C   I   I   Y   G   V   V   S   L   G   Y   Y   N   Q   C

154  AAT CTA TTA GTA AAA GTC TAT GCT GTC ATA TCC ATC AAA TCT TTA AAG CAG
       N   L   L   V   K   V   Y   A   V   I   S   I   K   S   L   K   Q

205  GTG TCT TTT GTA ACA AAT TCA ATA AAC ACC CAT TAA CTG TTT GTG CGC TAT
       V   S   F   V   T   N   S   I   N   T   H   -   L   F   V   R   Y

256  GGG CTA TAT TAG ATT GAA GCT TCT CAC ATG GCA TAT TTC AAA CTA ATA ATA
       G   L   Y   -   I   E   A   S   H   M   A   Y   F   K   L   I   I

307  TAA AAC CCT CAA GGA GGT GAA CTC ATA AGA AGG GTT GTA AAC CTT TTT TAA
       -   N   P   Q   G   G   E   L   I   R   R   V   V   N   L   F   -

358  TTA GTT TTT AAT TAG TTT GAA ATG CAA AGA AAT GAG AAA GCT TCA ATC CTG
       L   V   F   N   -   F   E   M   Q   R   N   E   K   A   S   I   L

409  ATT TAG CAA TTC TTT AAC GGG AGG CAA AGA ATG AAA TTT AGA AGG TCA ATT
       I   -   Q   F   F   N   G   R   Q   R   M   K   F   R   R   S   I

460  TGT ACT GCT GTT TTG TTG GCG GTT TTA TTG ACA CTT CTG GTA CCG ACA TCC
       C   T   A   V   L   L   A   V   L   L   T   L   L   V   P   T   S

511  GTG TTT GCC TTA GAA GAT AAT TCT TCG ACT TTG CCG CCG TAT AAA AAC GAC
       V   F   A   L   E   D   N   S   S   T   L   P   P   Y   K   N   D

562  CTT TTG TAT GAG AGG ACT TTT GAT GAG GGA CTT TGT TAT CCA TGG CAT ACC
       L   L   Y   E   R   T   F   D   E   G   L   C   Y   P   W   H   T

613  TGT GAA GAC AGC GGA GGA AAA TGC TCC TTT GAT GTG GTC GAT GTT CCG GGG
       C   E   D   S   G   G   K   C   S   F   D   V   V   D   V   P   G

664  CAG CCC GGT AAT AAA GCA TTT GCC GTT ACT GTT CTT GAC AAA GGG CAA AAC
       Q   P   G   N   K   A   F   A   V   T   V   L   D   K   G   Q   N

715  AGA TGG AGA GTT CAG ATG AGA CAC CGT GGT CTT ACT CTT GAA CAG GGA CAT
       R   W   R   V   Q   M   R   H   R   G   L   T   L   E   Q   G   H

766  ACA TAT AGA GTA CGG CTT AAG ATT TGG GCA GAT GCG TCC TGT AAA GTT TAT
       T   Y   R   V   R   L   K   I   W   A   D   A   S   C   K   V   Y

817  ATA AAA ATA GGA CAA ATG GCG GAG CCC TAT GCT GAA TAT TGG AAC AAC AAG
       I   K   I   G   Q   M   A   E   P   Y   A   E   Y   W   N   N   K

868  TGG AGT CCA TAC ACA CTG ACA GCA GGT AAG GTA TTG GAA ATT GAC GAG ACG
       W   S   P   Y   T   L   T   A   G   K   V   L   E   I   D   E   T

919  TTT GTT ATG GAC AAG CCA ACT GAC GAC ACA TGC GAA TTT ACA TTC CAT TTA
       F   V   M   D   K   P   T   D   D   T   C   E   F   T   F   H   L

970  GGT GGC GAA TTG GCA GCA ACT CCT CCA TAT ACA GTT TAT CTT GAT GAT GTA
       G   G   E   L   A   A   T   P   P   Y   T   V   Y   L   D   D   V
```

```
1021  TCC CTT TAT GAC CCA GAA TAT ACG AAG CCT GTT GAA TAT ATA CTT CCG CAG
        S   L   Y   D   P   E   Y   T   K   P   V   E   Y   I   L   P   Q

1072  CCT GAT GTA CGT GTG AAC CAG GTT GGC TAC CTG CCG GAG GGC AAG AAA GTT
        P   D   V   R   V   N   Q   V   G   Y   L   P   E   G   K   K   V

1123  GCC ACT GTG GTA TGC AAT TCA ACT CAG CCG GTA AAA TGG CAG CTT AAG AAT
        A   T   V   V   C   N   S   T   Q   P   V   K   W   Q   L   K   N

1174  GCT GCA GGC GTT GTA GTT TTG GAA GGT TAT ACC GAA CCA AAG GGT CTT GAC
        A   A   G   V   V   V   L   E   G   Y   T   E   P   K   G   L   D

1225  AAA GAC TCG CAG GAT TAT GTA CAT TGG CTT GAT TTT TCC GAT TTT GCA ACC
        K   D   S   Q   D   Y   V   H   W   L   D   F   S   D   F   A   T

1276  GAA GGA ATT GGT TAC TAT TTT GAA CTT CCG ACT GTA AAC AGT CCT ACA AAC
        E   G   I   G   Y   Y   F   E   L   P   T   V   N   S   P   T   N

1327  TAC AGT CAT CCA TTT GAC ATT CGC AAA GAC ATC TAT ACT CAG ATG AAA TAT
        Y   S   H   P   F   D   I   R   K   D   I   Y   T   Q   M   K   Y

1378  GAT GCA TTG GCA TTC TTC TAT CAC AAG AGA AGC GGT ATT CCT ATT GAA ATG
        D   A   L   A   F   F   Y   H   K   R   S   G   I   P   I   E   M

1429  CCG TAT GCA GGA GGA GAA CAG TGG ACC AGA CCT GCA GGA CAT ATC GGA ATT
        P   Y   A   G   G   E   Q   W   T   R   P   A   G   H   I   G   I

1480  GAG CCG AAC AAG GGA GAT ACA AAT GTT CCT ACA TGG CCT CAG GAT GAT GAG
        E   P   N   K   G   D   T   N   V   P   T   W   P   Q   D   D   E

1531  TAT GCA GGA ATA CCT CAG AAG AAT TAT ACA AAG GAT GTA ACC GGT GGA TGG
        Y   A   G   I   P   Q   K   N   Y   T   K   D   V   T   G   G   W

1582  TAT GAT GCC GGT GAC CAC GGT AAA TAT GTT GTA AAC GGC GGT ATA GCC GTC
        Y   D   A   G   D   H   G   K   Y   V   V   N   G   G   I   A   V

1633  TGG ACA TTA ATG AAC ATG TAT GAG AGG GCA AAA ATT AGA GGT CTT GAC AAC
        W   T   L   M   N   M   Y   E   R   A   K   I   R   G   L   D   N

1684  TGG GGA CCA TAC AGG GAC GGC GGA ATG AAC ATA CCG GAG CAG AAT AAC GGT
        W   G   P   Y   R   D   G   G   M   N   I   P   E   Q   N   N   G

1735  TAT CCG GAC ATT CTT GAT GAA GCA AGA TGG GAA ATT GAG TTC TTT AAG AAA
        Y   P   D   I   L   D   E   A   R   W   E   I   E   F   F   K   K

1786  ATG CAG GTA ACT GAA AAA GAG GAT CCT TCC ATA GCC GGA ATG GTA CAC CAC
        M   Q   V   T   E   K   E   D   P   S   I   A   G   M   V   H   H

1837  AAA ATT CAC GAC TTC AGA TGG ACT GCT TTG GGT ATG TTG CCT CAC GAA GAT
        K   I   H   D   F   R   W   T   A   L   G   M   L   P   H   E   D

1888  CCC CAG CCA CGT TAC TTA AGG CCG GTA AGT ACG GCT GCG ACT TTG AAC TTT
        P   Q   P   R   Y   L   R   P   V   S   T   A   A   T   L   N   F

1939  GCG GCA ACT TTG GCA CAA AGT GCA CGT CTT TGG AAA GAT TAT GAT CCG ACT
        A   A   T   L   A   Q   S   A   R   L   W   K   D   Y   D   P   T

1990  TTT GCT GCT GAC TGT TTG GAA AAG GCT GAA ATA GCA TGG CAG GCG GCA TTA
        F   A   A   D   C   L   E   K   A   E   I   A   W   Q   A   A   L

2041  AAG CAT CCT GAT ATT TAT GCT GAG TAT ACT CCC GGT AGC GGT GGT CCC GGA
        K   H   P   D   I   Y   A   E   Y   T   P   G   S   G   G   P   G
```

```
2092  GGC GGA CCA TAC AAT GAC GAC TAT GTC GGA GAC GAA TTC TAC TGG GCA GCC
        G   G   P   Y   N   D   D   Y   V   G   D   E   F   Y   W   A   A

2143  TGC GAA CTT TAT GTA ACA ACA GGA AAA GAC GAA TAT AAG AAT TAC CTG ATG
        C   E   L   Y   V   T   T   G   K   D   E   Y   K   N   Y   L   M

2194  AAT TCA CCT CAC TAT CTT GAA ATG CCT GCA AAG ATG GGT GAA AAC GGT GGA
        N   S   P   H   Y   L   E   M   P   A   K   M   G   E   N   G   G

2245  GCA AAC GGA GAA GAC AAC GGA TTG TGG GGA TGC TTC ACC TGG GGA ACT ACT
        A   N   G   E   D   N   G   L   W   G   C   F   T   W   G   T   T

2296  CAA GGA TTG GGA ACC ATT ACT CTT GCT TTG GTT GAA AAC GGA TTG CCT GCT
        Q   G   L   G   T   I   T   L   A   L   V   E   N   G   L   P   A

2347  ACA GAC ATT CAA AAG GCA AGA AAC AAT ATA GCT AAA GCT GCT GAC AGA TGG
        T   D   I   Q   K   A   R   N   N   I   A   K   A   A   D   R   W

2398  CTT GAG AAT ATT GAA GAG CAA GGT TAC AGA CTG CCG ATC AAA CGG GCG GAG
        L   E   N   I   E   E   Q   G   Y   R   L   P   I   K   R   A   E

2449  GAT GAG AGA GCC GGT TAT CCA TGG GGT TCA AAC TCC TTG CAT TTT GAA CCA
        D   E   R   A   G   Y   P   W   G   S   N   S   L   H   F   E   P

2500  GAT GAC CTA GTT ATG GGA TAT GCC TAT GAC TTT ACA GGT GAC TCA AAT ATC
        D   D   L   V   M   G   Y   A   Y   D   F   T   G   D   S   N   I

2551  TCG ATG GAA TGT TTG ACC GGC ATA AGC TAC CTG TTG GGA AGA AAC GCA ATG
        S   M   E   C   L   T   G   I   S   Y   L   L   G   R   N   A   M

2602  GAT CAG TCC TAT GTA ACA GGG TAT GGT GAG CGT CCG CTT CAG AAT CCT CAT
        D   Q   S   Y   V   T   G   Y   G   E   R   P   L   Q   N   P   H

2653  GAC AGG TTC TGG ACG CCG CAG ACA AGT AAG AGA TTC CCT GCT CCA CCT CCG
        D   R   F   W   T   P   Q   T   S   K   R   F   P   A   P   P   P

2704  GGT ATA ATT TCC GGC CGT CCG AAC TCC CGT TTC GAG GAC CCG ACA ATA AAT
        G   I   I   S   G   R   P   N   S   R   F   E   D   P   T   I   N

2755  GCG GCC GTT AAG AAG GAT ACA CCG CCA CAG AAA TGT TTT ATC GAC CAT ACA
        A   A   V   K   K   D   T   P   P   Q   K   C   F   I   D   H   T

2806  GAC TCA TGG TCA ACC AAC GAG ATA ACT GTT AAC TGG AAT GCT CCG TTT GCA
        D   S   W   S   T   N   E   I   T   V   N   W   N   A   P   F   A

2857  TGG GTT ACA GCT TAT CTT GAC GAG CAG TAC ACA GAC AGT GAA ACC GAT AAG
        W   V   T   A   Y   L   D   E   Q   Y   T   D   S   E   T   D   K

2908  GTA ACT ATT GAT TCG CCT GTT GCA GGA GAA AGA TTT GAA GCG GGT AAA GAC
        V   T   I   D   S   P   V   A   G   E   R   F   E   A   G   K   D

2959  ATT AAT ATA AGA ACT GTT AAA TCA AAA ACT CCT GTA AGC AAA GTA GAG TTT
        I   N   I   R   T   V   K   S   K   T   P   V   S   K   V   E   F

3010  TAC AAT GGA GAT ACG CTT ATT TCC AGT GAC ACA ACT GCA CCT TAC ACA GCA
        Y   N   G   D   T   L   I   S   S   D   T   T   A   P   Y   T   A

3061  AAG ATA ACA GGA GCC GCT GTC GGA GCA TAT AAC CTT AAA GCG GTT GCA GTG
        K   I   T   G   A   A   V   G   A   Y   N   L   K   A   V   A   V

3112  CTG TCT GAC GGA AGA AGA ATT GAG TCA CCG GTA ACT CCT GTA CTT GTT AAG
        L   S   D   G   R   R   I   E   S   P   V   T   P   V   L   V   K
```

```
3163  GTA ATT GTG AAA CCT ACT GTA AAA CTT ACT GCA CCC AAG TCA AAT GTT GTG
        V   I   V   K   P   T   V   K   L   T   A   P   K   S   N   V   V

3214  GCT TAT GGA AAT GAG TTC CTG AAG ATT ACA GCA ACA GCC AGT GAC TCT GAC
        A   Y   G   N   E   F   L   K   I   T   A   T   A   S   D   S   D

3265  GGC AAA ATC TCC AGG GTT GAT TTC CTT GTT GAC GGT GAA GTA ATC GGT TCA
        G   K   I   S   R   V   D   F   L   V   D   G   E   V   I   G   S

3316  GAC AGG GAA GCA CCT TAT GAA TAT GAG TGG AAA GCT GTG GAA GGC AAT CAC
        D   R   E   A   P   Y   E   Y   E   W   K   A   V   E   G   N   H

3367  GAA ATA AGT GTA ATT GCT TAT GAT GAT GAC GAT GCG GCT TCA ACA CCT GAT
        E   I   S   V   I   A   Y   D   D   D   D   A   A   S   T   P   D

3418  TCC GTA AAA ATA TTT GTA AAA CAG GCA CGG GAT GTA AAA GTA CAG TAT TTG
        S   V   K   I   F   V   K   Q   A   R   D   V   K   V   Q   Y   L

3469  TGC GAA AAT ACG CAA ACA TCC ACT CAG GAA ATC AAG GGT AAA TTC AAT ATA
        C   E   N   T   Q   T   S   T   Q   E   I   K   G   K   F   N   I

3520  GTT AAC ACA GGA AAC AGA GAT TAT TCG CTG AAA GAT ATA GTA TTA AGA TAC
        V   N   T   G   N   R   D   Y   S   L   K   D   I   V   L   R   Y

3571  TAC TTT ACC AAG GAG CAC AAT TCA CAG CTT CAG TTT ATC TGC TAT TAT ACA
        Y   F   T   K   E   H   N   S   Q   L   Q   F   I   C   Y   Y   T

3622  CCC ATA GGC TCC GGA AAT CTC ATT CCG TCC TTT GGC GGC TCG GGT GAC GAG
        P   I   G   S   G   N   L   I   P   S   F   G   G   S   G   D   E

3673  CAT TAT CTG CAG CTG GAA TTC AAA GAT GTC AAG CTG CCT GCC GGC GGT CAG
        H   Y   L   Q   L   E   F   K   D   V   K   L   P   A   G   G   Q

3724  ACT GGG GAA ATA CAG TTT GTT ATA AGA TAT GCA GAT AAC TCC TTC CAT GAT
        T   G   E   I   Q   F   V   I   R   Y   A   D   N   S   F   H   D

3775  CAG TCG AAC GAC TAT TCG TTC GAT CCA ACT ATA AAA GCG TTC CAG GAT TAT
        Q   S   N   D   Y   S   F   D   P   T   I   K   A   F   Q   D   Y

3826  GGC AAG GTT ACC CTG TAT AAG AAT GGA GAA CTT GTT TGG GGA ACG CCG CCG
        G   K   V   T   L   Y   K   N   G   E   L   V   W   G   T   P   P

3877  GGC GGT ACA GAA CCT GAA GAA CCG GAA GAG CCT GAA GAA CCG GAA GAG CCT
        G   G   T   E   P   E   E   P   E   E   P   E   E   P   E   E   P

3928  GCG ATA GTT TAC GGC GAC TGT AAT GAT GAC GGC AAA GTA AAT TCA ACA GAC
        A   I   V   Y   G   D   C   N   D   D   G   K   V   N   S   T   D

3979  GTC GCA GTA ATG AAG AGA TAT TTA AAG AAA GAA AAT GTT AAT ATT AAT CTT
        V   A   V   M   K   R   Y   L   K   K   E   N   V   N   I   N   L

4030  GAC AAT GCA GAT GTG AAT GCG GAC GGC AAA GTT AAC TCA ACA GAC TTC TCA
        D   N   A   D   V   N   A   D   G   K   V   N   S   T   D   F   S
                                                                        •
4081  ATA CTT AAG AGA TAT GTT ATG AAG AAC ATA GAA GAA TTG CCA TAT CGA TAA
        I   L   K   R   Y   V   M   K   N   I   E   E   L   P   Y   R   -

4132  GAT AAT CTG AAA TTA TTT GTG TAA GGA CCG GTT TTT GCC GGT CCT TTT TAA
        D   N   L   K   L   F   V   -   G   P   V   F   A   G   P   F   -
```

Endoglucanase J (GH9)

```
  1  TTT ATA AAT ATT TAT AAA AAA CAT GTA AGA TTG ATA AAT GAA GGC TTT TTA
       F   I   N   I   Y   K   K   H   V   R   L   I   N   E   G   F   L

 52  TAT TAT TTT TTG TAA TTA TTG CAT TTT GCC CCC TTA AAA AAT TTT AAA ATT
       Y   Y   F   L   -   L   L   H   F   A   P   L   K   N   F   K   I

103  CGA AAT TAA TTT TTG CGT ATG GAT TGT GCA ATA AAA ATA TGA TAA TAT TTA
       R   N   -   F   L   R   M   D   C   A   I   K   I   -   -   Y   L

154  GGT GAA TAT AAA AAG ATT GTA TTA ATC AAT TGC TTA CAG GTA TAA AAA TTA
       G   E   Y   K   K   I   V   L   I   N   C   L   Q   V   -   K   L

205  AGA AAA TTG CGC ACT GTG TTG GCG TAA CCT TAA TCA ACA ATT GGA AAT GGG
       R   K   L   R   T   V   L   A   -   P   -   S   T   I   G   N   G

256  GTT TGG GAA GTA TAC TTG GAA GCA ACT GGT TAT TAT TGA AAT TTA CAG ACG
       V   W   E   V   Y   L   E   A   T   G   Y   Y   -   N   L   Q   T

307  ATG TGA AAG ACT GTG AAC AAA CCA ATC AAT GCC CAA ATT ATA CTT TGC AGC
       M   -   K   T   V   N   K   P   I   N   A   Q   I   I   L   C   S

358  AAA ATC ATC CTT GCA TAT CCC TTC AGA TTT AAA AGC CGC GAC AAT TAG CGT
       K   I   I   L   A   Y   P   F   R   F   K   S   R   D   N   -   R

409  TTC CGA TGA AGC TCA TTT ATG TGA ACC TCG ATT CCT TCC TAT AAA ATT TAC
       F   R   -   S   S   F   M   -   T   S   I   P   S   Y   K   I   Y

460  TCT ATT TTC ATT TCA ATT TTT CCA GTT TGC ATG TGG AAT TAT TTC CCG AAG
       S   I   F   I   S   I   F   P   V   C   M   W   N   Y   F   P   K

511  CTC CGT ATT TCT GTT GTT GAT GGT TTG TAT GTC AAC ATA TGC GTG CAT TTT
       L   R   I   S   V   V   D   G   L   Y   V   N   I   C   V   H   F

562  AAT ATA AAA TGA ATT GTT AAA GGA GTG TGA AAT ATG GCA AAG AGA AGA TTA
       N   I   K   -   I   V   K   G   V   -   N   M   A   K   R   R   L

613  TCG CTA CTT TTG GTA CTT GCC ATA ATG TTT ACG ATG GTC GTT CCA CAG ATA
       S   L   L   L   V   L   A   I   M   F   T   M   V   V   P   Q   I

664  TCT GCA AGT GCC GAA ACA GTT GCT CCT GAA GGC TAC AGG AAG CTT TTG GAT
       S   A   S   A   E   T   V   A   P   E   G   Y   R   K   L   L   D

715  GTA CAA ATT TTC AAG GAT TCG CCT GTA GTC GGA TGG TCA GGA AGC GGT ATG
       V   Q   I   F   K   D   S   P   V   V   G   W   S   G   S   G   M

766  GGC GAG CTT GAA ACT ATC GGC GAT ACC CTT CCG GTT GAT ACC ACA GTT ACA
       G   E   L   E   T   I   G   D   T   L   P   V   D   T   T   V   T

817  TAT AAC GGT TTG CCG ACT TTA AGA CTG AAT GTC CAG ACA ACC GTT CAG TCA
       Y   N   G   L   P   T   L   R   L   N   V   Q   T   T   V   Q   S

868  GGA TGG TGG ATT TCT CTT CTT ACA TTA AGA GGA TGG AAC ACC CAT GAC CTT
       G   W   W   I   S   L   L   T   L   R   G   W   N   T   H   D   L

919  TCC CAG TAT GTC GAA AAC GGT TAT CTT GAG TTT GAC ATC AAG GGT AAG GAA
       S   Q   Y   V   E   N   G   Y   L   E   F   D   I   K   G   K   E

970  GGC GGA GAA GAC TTT GTT ATT GGT TTC AGG GAC AAG GTT TAT GAA CGC GTT
       G   G   E   D   F   V   I   G   F   R   D   K   V   Y   E   R   V
```

```
1021  TAC GGA CTT GAA ATT GAT GTT ACC ACA GTA ATA TCA AAT TAT GTA ACG GTA
        Y   G   L   E   I   D   V   T   T   V   I   S   N   Y   V   T   V

1072  ACT ACG GAC TGG CAG CAT GTT AAG ATT CCT TTG AGA GAC CTG ATG AAG ATT
        T   T   D   W   Q   H   V   K   I   P   L   R   D   L   M   K   I

1123  AAT AAC GGA TTT GAT CCT TCA TCA GTT ACA TGC CTG GTG TTC TCA AAA AGA
        N   N   G   F   D   P   S   S   V   T   C   L   V   F   S   K   R

1174  TAT GCA GAT CCG TTT ACA GTA TGG TTC AGT GAT ATA AAG ATT ACA TCA GAA
        Y   A   D   P   F   T   V   W   F   S   D   I   K   I   T   S   E

1225  GAC AAT GAA AAG TCC GCT CCT GCA ATC AAG GTA AAC CAG CTT GGC TTT ATT
        D   N   E   K   S   A   P   A   I   K   V   N   Q   L   G   F   I

1276  CCT GAA GCT GAA AAA TAC GCT TTG GTT ACA GGT TTT GCA GAA GAG CTC GCA
        P   E   A   E   K   Y   A   L   V   T   G   F   A   E   E   L   A

1327  GTA TCG GAA GGT GAC GAA TTT GCC GTT ATA AAT GCT GCG GAC AAT TCT GTT
        V   S   E   G   D   E   F   A   V   I   N   A   A   D   N   S   V

1378  GCT TAT ACC GGA AAA TTA ACT CTT GTA ACA GAA TAT GAA CCT CTT GAT TCC
        A   Y   T   G   K   L   T   L   V   T   E   Y   E   P   L   D   S

1429  GGA GAA AAA ATA CTT AAG GCA GAT TTC AGC GAC TTG ACT GTA CCT GGC AAA
        G   E   K   I   L   K   A   D   F   S   D   L   T   V   P   G   K

1480  TAC TAC ATT AGT ATT GAA GGT CTT GAC AAT TCA CCC AAG TTT GAA ATC GGT
        Y   Y   I   S   I   E   G   L   D   N   S   P   K   F   E   I   G

1531  GAA GGT ATT TAC GGT CCA CTG GTT GTT GAC GCT GCA AGA TAT TTC TAT TAT
        E   G   I   Y   G   P   L   V   V   D   A   A   R   Y   F   Y   Y

1582  CAG CGT CAG GGT ATA GAA CTT GAA GAG CCT TAT GCG CAG GGA TAT CCC CGC
        Q   R   Q   G   I   E   L   E   E   P   Y   A   Q   G   Y   P   R

1633  AAG GAC GTT ACT CCT CAG GAC GCA TAT GCT GTA TTT GCA TCC GGA AAG AAG
        K   D   V   T   P   Q   D   A   Y   A   V   F   A   S   G   K   K

1684  GAT CCG ATT GAC ATA ACA AAG GGT TGG TAT GAC GCA GGA GAC TTC GGT AAG
        D   P   I   D   I   T   K   G   W   Y   D   A   G   D   F   G   K

1735  TAT GTA AAT GCC GGA GCA ACC GGT GTT TCC GAT TTG TTC TGG GCA TAT GAA
        Y   V   N   A   G   A   T   G   V   S   D   L   F   W   A   Y   E

1786  ATG TTC CCT TCC CAG TTT GTT GAC GGT CAG TTC AAT ATT CCT GAA AGC GGA
        M   F   P   S   Q   F   V   D   G   Q   F   N   I   P   E   S   G

1837  AAC GGT GTA CCG GAC ATC CTT GAC GAA GCT CGC TGG GAG CTT GAA TGG ATG
        N   G   V   P   D   I   L   D   E   A   R   W   E   L   E   W   M

1888  CTG AAA ATG CAG GAC AAA GAA AGC GGA GGA TTC TAT CCC AGA GTT CAA TCT
        L   K   M   Q   D   K   E   S   G   G   F   Y   P   R   V   Q   S

1939  GAC AAT GAC GAA AAC ATA AAA TCA AGA ATA ATC AGG GAT CAG AAC GGC TGT
        D   N   D   E   N   I   K   S   R   I   I   R   D   Q   N   G   C

1990  ACC ACT GAT GAT ACT GCA TGT GCC GCC GGA ATA CTT GCT CAT GCA TAC TTG
        T   T   D   D   T   A   C   A   A   G   I   L   A   H   A   Y   L

2041  ATT TAC AAG GAT ATT GAC CCT GAT TTT GCA CAA GAG TGC CTG GAT GCG GCA
        I   Y   K   D   I   D   P   D   F   A   Q   E   C   L   D   A   A
```

```
2092  ATA AAT GCA TGG AAA TTC CTT GAA AAG AAT CCT GAA AAC ATT GTT TCA CCT
        I   N   A   W   K   F   L   E   K   N   P   E   N   I   V   S   P

2143  CCG GGT CCA TAC AAC GTA TAT GAC GAC AGC GGA GAC AGA CTC TGG GCT GCA
        P   G   P   Y   N   V   Y   D   D   S   G   D   R   L   W   A   A

2194  GCT TCG CTG TAC AGA GCT ACC GGT GAA GAG GTT TAT CAT ACA TAC TTT AAA
        A   S   L   Y   R   A   T   G   E   E   V   Y   H   T   Y   F   K

2245  CAA AAC TAC AAA TCT TTT GCA CAA AAG TTC GAA AGC CCG ACT GCA TAT GCT
        Q   N   Y   K   S   F   A   Q   K   F   E   S   P   T   A   Y   A

2296  CAT ACA TGG GGT GAT ATG TGG CTT ACG GCA TTC CTT TCG TAT TTG AAA GCT
        H   T   W   G   D   M   W   L   T   A   F   L   S   Y   L   K   A

2347  GAA AAC AAG GAT CAG GAA GTT GTA GAC TGG ATT GAT ACA GAG TTT GGA ATC
        E   N   K   D   Q   E   V   V   D   W   I   D   T   E   F   G   I

2398  TGG CTT GAA AAC ATA CTC ACA AGA TAT GAG AAC AAT CCA TGG AAG AAT GCA
        W   L   E   N   I   L   T   R   Y   E   N   N   P   W   K   N   A

2449  ATT GTT CCC GGA AAC TAC TTC TGG GGA ATC AAC ATG CAG GTT ATG AAT GTT
        I   V   P   G   N   Y   F   W   G   I   N   M   Q   V   M   N   V

2500  CCG ATG GAT GCT ATC ATA GGT TCA CAG CTT CTT GGA AAA TAC AGT GAC AGA
        P   M   D   A   I   I   G   S   Q   L   L   G   K   Y   S   D   R

2551  ATA GAA AAA TTA GGT TTT GGT TCA CTT AAC TGG CTG CTT GGT ACA AAT CCG
        I   E   K   L   G   F   G   S   L   N   W   L   L   G   T   N   P

2602  CTT CGC TTC AGC TTT GTA TCA GGA TAT GGA GAG GAT TCT GTA AAA GGA GTA
        L   R   F   S   F   V   S   G   Y   G   E   D   S   V   K   G   V

2653  TTC AGC AAT ATT TAC AAT ACG GAC GGC AAG CAG GGA ATT CCG AAA GGA TAC
        F   S   N   I   Y   N   T   D   G   K   Q   G   I   P   K   G   Y

2704  ATG CCT GGT GGA CCA AAT GCT TAT GAA GGT GCA GGC CTG TCA AGG TTT GCA
        M   P   G   G   P   N   A   Y   E   G   A   G   L   S   R   F   A

2755  GCA AAA TGC TAC ACC AGA AGT ACC GGT GAC TGG GTA GCC AAC GAA CAT ACA
        A   K   C   Y   T   R   S   T   G   D   W   V   A   N   E   H   T

2806  GTA TAT TGG AAC TCA GCT TTG GTA TTT ATG GCT GCT TTT GCA AAC CAG GGT
        V   Y   W   N   S   A   L   V   F   M   A   A   F   A   N   Q   G

2857  TCA GAG GTT AAT CCG GGA CCT GCG CCG GAA CCG GGA GTA ACT CCG AAT CCT
        S   E   V   N   P   G   P   A   P   E   P   G   V   T   P   N   P

2908  ACA GAA CCT GCA AAA GTG GTT GAC ATC AGG ATA GAT ACT TCT GCT GAA AGA
        T   E   P   A   K   V   V   D   I   R   I   D   T   S   A   E   R

2959  AAG CCA ATC AGC CCG TAT ATA TAC GGA AGC AAT CAG GAA CTT GAT GCA ACA
        K   P   I   S   P   Y   I   Y   G   S   N   Q   E   L   D   A   T

3010  GTT ACT GCA AAG AGG TTC GGC GGA AAC AGA ACT ACA GGA TAC AAC TGG GAA
        V   T   A   K   R   F   G   G   N   R   T   T   G   Y   N   W   E

3061  AAC AAC TTC TCA AAT GCA GGA AGT GAC TGG CTG CAT TAC AGT GAT ACA TAC
        N   N   F   S   N   A   G   S   D   W   L   H   Y   S   D   T   Y

3112  CTT TTG GAG GAC GGC GGA GTT CCT AAG GGA GAG TGG AGT ACA CCT GCT TCT
        L   L   E   D   G   G   V   P   K   G   E   W   S   T   P   A   S
```

```
3163  GTA GTT ACC ACG TTC CAT GAC AAG GCA CTT AGC AAA AAT GTT CCT TAC ACA
        V   V   T   T   F   H   D   K   A   L   S   K   N   V   P   Y   T

3214  CTT ATC ACT CTT CAG GCA GCA GGT TAT GTT TCC GCA GAC GGA AAC GGA CCG
        L   I   T   L   Q   A   A   G   Y   V   S   A   D   G   N   G   P

3265  GTT TCC CAG GAA GAA ACT GCA CCG TCT TCA AGA TGG AAG GAA GTT AAG TTT
        V   S   Q   E   E   T   A   P   S   S   R   W   K   E   V   K   F

3316  GAA AAG GGA GCA CCT TTC TCA CTT ACA CCG GAC ACA GAA GAT GAT TAT GTT
        E   K   G   A   P   F   S   L   T   P   D   T   E   D   D   Y   V

3367  TAC ATG GAT GAG TTT GTA AAC TAT CTT GTA AAC AAA TAC GGA AAT GCA TCC
        Y   M   D   E   F   V   N   Y   L   V   N   K   Y   G   N   A   S

3418  ACA CCT ACA GGA ATA AAG GGT TAT TCA ATA GAT AAC GAG CCG GCA TTG TGG
        T   P   T   G   I   K   G   Y   S   I   D   N   E   P   A   L   W

3469  AGT CAT ACT CAT CCG AGA ATT CAT CCG GAC AAT GTA ACT GCC AAA GAG CTT
        S   H   T   H   P   R   I   H   P   D   N   V   T   A   K   E   L

3520  ATT GAA AAA TCT GTA GCT CTT TCC AAG GCG GTT AAA AAG GTA GAT CCA TAT
        I   E   K   S   V   A   L   S   K   A   V   K   K   V   D   P   Y

3571  GCA GAA ATA TTC GGA CCT GCT TTG TAC GGA TTT GCC GCA TAT GAG ACA CTT
        A   E   I   F   G   P   A   L   Y   G   F   A   A   Y   E   T   L

3622  CAG TCA GCT CCT GAC TGG GGA ACT GAA GGA GAA GGA TAC AGG TGG TTT ATA
        Q   S   A   P   D   W   G   T   E   G   E   G   Y   R   W   F   I

3673  GAT TAT TAC CTC GAT AAG ATG AAA AAG GCT TCT GAT GAA GAA GGA AAG AGA
        D   Y   Y   L   D   K   M   K   K   A   S   D   E   E   G   K   R

3724  CTT TTG GAC GTA CTT GAC GTA CAC TGG TAT CCG GAA GCC AGG GGC GGC GGT
        L   L   D   V   L   D   V   H   W   Y   P   E   A   R   G   G   G

3775  GAA AGA ATA TGC TTT GGA GCC GAT CCA AGA AAT ATT GAG ACA AAC AAA GCA
        E   R   I   C   F   G   A   D   P   R   N   I   E   T   N   K   A

3826  AGA TTG CAG GCG CCC AGA ACA TTG TGG GAT CCT ACA TAT ATT GAA GAC AGC
        R   L   Q   A   P   R   T   L   W   D   P   T   Y   I   E   D   S

3877  TGG ATA GGA CAA TGG AAG AAG GAT TTC CTC CCG ATA TTA CCT AAT CTT TTG
        W   I   G   Q   W   K   K   D   F   L   P   I   L   P   N   L   L

3928  GAT TCC ATT GAA AAA TAT TAT CCG GGA ACG AAG CTT GCT ATA ACT GAA TAT
        D   S   I   E   K   Y   Y   P   G   T   K   L   A   I   T   E   Y

3979  GAC TAT GGC GGA GGA AAT CAT ATT ACA GGC GGT ATT GCT CAA GCC GAT GTT
        D   Y   G   G   G   N   H   I   T   G   G   I   A   Q   A   D   V

4030  CTT GGT ATA TTC GGT AAA TAC GGT GTT TAC CTT GCA ACA TTC TGG GGA GAT
        L   G   I   F   G   K   Y   G   V   Y   L   A   T   F   W   G   D

4081  GCA AGC AAT AAC TAT ACT GAG GCC GGT ATA AAC CTT TAT ACC AAC TAC GAC
        A   S   N   N   Y   T   E   A   G   I   N   L   Y   T   N   Y   D

4132  GGC AAA GGC GGC AAA TTT GGA GAT ACA TCC GTA AAA TGT GAA ACG TCC GAC
        G   K   G   G   K   F   G   D   T   S   V   K   C   E   T   S   D

4183  ATA GAA GTA AGC TCT GCT TAT GCA TCC ATT GTC GGT GAA GAT GAC AGC AAA
        I   E   V   S   S   A   Y   A   S   I   V   G   E   D   D   S   K
```

```
4234  CTC CAT ATC ATT CTT TTG AAC AAG AAC TAT GAC CAG CCG ACG ACA TTC AAT
        L   H   I   I   L   L   N   K   N   Y   D   Q   P   T   T   F   N

4285  TTC TCA ATT GAC AGC AGC AAG AAC TAC ACA ATA GGA AAT GTA TGG GCA TTT
        F   S   I   D   S   S   K   N   Y   T   I   G   N   V   W   A   F

4336  GAC AGA GGA AGC TCC AAT ATT ACT CAA AGA ACT CCT ATA GTG AAC ATA AAG
        D   R   G   S   S   N   I   T   Q   R   T   P   I   V   N   I   K

4387  GAC AAT ACC TTC ACA TAT ACA GTA CCG GCT TTG ACA GCG TGC CAT ATT GTG
        D   N   T   F   T   Y   T   V   P   A   L   T   A   C   H   I   V

4438  CTT GAA GCT GCG GAG CCC GTA GTG TAC GGA GAC TTG AAC AAT GAC TCT AAA
        L   E   A   A   E   P   V   V   Y   G   D   L   N   N   D   S   K

4489  GTA AAC GCA GTA GAC ATT ATG ATG CTC AAA CGA TAT ATT CTC GGA ATA ATA
        V   N   A   V   D   I   M   M   L   K   R   Y   I   L   G   I   I

4540  GAT AAT ATA AAT CTG ACA GCA GCT GAC ATT TAT TTT GAC GGT GTT GTA AAT
        D   N   I   N   L   T   A   A   D   I   Y   F   D   G   V   V   N

4591  TCA AGT GAC TAT AAT ATA ATG AAG AGA TAT TTG TTA AAG GCA ATA GAA GAT
        S   S   D   Y   N   I   M   K   R   Y   L   L   K   A   I   E   D

4642  ATT CCT TAT GTT CCG GAA AAC CAG GCA CCT AAA GCA ATA TTT ACT TTC TCG
        I   P   Y   V   P   E   N   Q   A   P   K   A   I   F   T   F   S

4693  CCC GAA GAC CCG GTT ACT GAC GAG AAT GTA GTG TTC AAT GCA TCA AAT TCA
        P   E   D   P   V   T   D   E   N   V   V   F   N   A   S   N   S

4744  ATA GAT GAA GAC GGA ACA ATT GCC TAT TAT GCA TGG GAT TTC GGT GAC GGA
        I   D   E   D   G   T   I   A   Y   Y   A   W   D   F   G   D   G

4795  TAT GAA GGA ACT TCA ACA ACA CCG ACT ATT ACC TAT AAG TAT AAA AAC CCC
        Y   E   G   T   S   T   T   P   T   I   T   Y   K   Y   K   N   P

4846  GGA ACA TAC AAA GTA AAA CTG ATT GTT ACA GAC AAC CAG GGG GCT TCA AGT
        G   T   Y   K   V   K   L   I   V   T   D   N   Q   G   A   S   S

4897  TCG TTT ACA GCT ACC ATA AAA GTA ACC TCA GCT ACC GGG GAC AAT TCC AAA
        S   F   T   A   T   I   K   V   T   S   A   T   G   D   N   S   K

4948  TTC AAC TTT GAA GAC GGC ACG CTG GGA GGA TTT ACA ACA TCC GGA ACA AAT
        F   N   F   E   D   G   T   L   G   G   F   T   T   S   G   T   N

4999  GCT ACG GGT GTT GTT GTG AAC ACT ACT GAA AAA GCA TTC AAA GGC GAA AGA
        A   T   G   V   V   V   N   T   T   E   K   A   F   K   G   E   R

5050  GGT CTT AAA TGG ACT GTA ACA AGC GAA GGA GAA GGA ACT GCA GAA TTG AAA
        G   L   K   W   T   V   T   S   E   G   E   G   T   A   E   L   K

5101  CTT GAC GGA GGT ACT ATT GTA GTT CCC GGT ACC ACT ATG ACG TTT AGA ATC
        L   D   G   G   T   I   V   V   P   G   T   T   M   T   F   R   I

5152  TGG ATA CCT TCC GGT GCG CCT ATT GCT GCC ATC CAG CCG TAT ATT ATG CCT
        W   I   P   S   G   A   P   I   A   A   I   Q   P   Y   I   M   P

5203  CAT ACA CCT GAT TGG TCG GAA GTC CTC TGG AAT TCG ACA TGG AAA GGA TAC
        H   T   P   D   W   S   E   V   L   W   N   S   T   W   K   G   Y

5254  ACC ATG GTG AAG ACC GAT GAC TGG AAT GAA ATT ACC CTG ACA CTG CCG GAA
        T   M   V   K   T   D   D   W   N   E   I   T   L   T   L   P   E
```

```
5305  GAC GTG GAT CCG ACT TGG CCG CAG CAG ATG GGT ATA CAG GTA CAG ACC ATA
        D   V   D   P   T   W   P   Q   Q   M   G   I   Q   V   Q   T   I
                                                                    ●
5356  GAT GAA GGT GAA TTC ACT ATC TAT GTA GAT GCT ATT GAC TGG TAA GAA ATT
        D   E   G   E   F   T   I   Y   V   D   A   I   D   W   -   E   I

5407  TAA TTG GTC AAA CAA AAA TCA AAT CAG TTG CAA GTT ATA AGG CCG GGC AAA
        -   L   V   K   Q   K   S   N   Q   L   Q   V   I   R   P   G   K

5458  AAA GGA GGA ACC TTT GCG GTT TCT CCT TTT TTG CTT TTA AAA TTT TTT TAT
        K   G   G   T   F   A   V   S   P   F   L   L   L   K   F   F   Y

5509  CGA GAT TAG TTC AAA AAA CAA AAA CGG GAA ACA GGA TTT GAC CCA CTT AAA
        R   D   -   F   K   K   Q   K   R   E   T   G   F   D   P   L   K
```

Exoglucanase K (GH9)

```
   1  TTG TTT TAA GTT GTT AAA GTG AAG CTT AAA GCT TTT CAA ACT AAT AAA TAA
        L   F   -   V   V   K   V   K   L   K   A   F   Q   T   N   K   -

  52  AAA CAC TGC AAA GGA GGT GTA CTC TTA TCA GAG GGC TGT AAA ACC CTG TAT
        K   H   C   K   G   G   V   L   L   S   E   G   C   K   T   L   Y

 103  TTA TTA GTT TGA TGG TAG GTT TAA TAA TTT GTT GCA TTT TTT ATG ATA TTC
        L   L   V   -   W   -   V   -   -   F   V   A   F   F   M   I   F

 154  GAC AGA AAG AAA AAA GAG TAT TGT AAA GAA TGC AAC GGA AAA GCT TCA TTT
        D   R   K   K   K   E   Y   C   K   E   C   N   G   K   A   S   F

 205  TCT AAT GCT TTA AAT TAA TTA ACG GGA GGT AGA TTT ATG AAT TTC AGA AGA
        S   N   A   L   N   -   L   T   G   G   R   F   M   N   F   R   R

 256  ATG TTG TGC GCA GCC ATA GTG TTG ACA ATT GTA CTG TCC ATT ATG CTG CCG
        M   L   C   A   A   I   V   L   T   I   V   L   S   I   M   L   P

 307  TCA ACT GTT TTT GCT TTG GAA GAC AAG TCT TCA AAG TTG CCA GAT TAT AAA
        S   T   V   F   A   L   E   D   K   S   S   K   L   P   D   Y   K

 358  AAC GAC CTT TTG TAT GAA AGA ACA TTC GAC GAA GGT CTT TGC TTT CCG TGG
        N   D   L   L   Y   E   R   T   F   D   E   G   L   C   F   P   W

 409  CAT ACT TGC GAA GAC AGT GGA GGA AAA TGT GAT TTC GCT GTT GTT GAT GTT
        H   T   C   E   D   S   G   G   K   C   D   F   A   V   V   D   V

 460  CCA GGA GAG CCT GGG AAC AAA GCT TTC CGC TTG ACA GTA ATT GAC AAA GGA
        P   G   E   P   G   N   K   A   F   R   L   T   V   I   D   K   G

 511  CAA AAC AAG TGG AGT GTC CAG ATG AGA CAC AGA GGT ATT ACC CTC GAG CAA
        Q   N   K   W   S   V   Q   M   R   H   R   G   I   T   L   E   Q

 562  GGA CAT ACA TAC ACG GTA AGG TTT ACG ATT TGG TCT GAC AAA TCC TGT AGG
        G   H   T   Y   T   V   R   F   T   I   W   S   D   K   S   C   R

 613  GTT TAT GCT AAA ATT GGT CAG ATG GGT GAA CCC TAT ACT GAA TAT TGG AAC
        V   Y   A   K   I   G   Q   M   G   E   P   Y   T   E   Y   W   N
```

```
 664  AAT AAC TGG AAT CCA TTC AAC CTT ACA CCA GGA CAG AAG CTT ACA GTT GAA
        N   N   W   N   P   F   N   L   T   P   G   Q   K   L   T   V   E

 715  CAG AAT TTT ACA ATG AAC TAT CCT ACT GAT GAC ACA TGC GAG TTC ACA TTC
        Q   N   F   T   M   N   Y   P   T   D   D   T   C   E   F   T   F

 766  CAT TTG GGT GGA GAA CTT GCT GCA GGT ACA CCT TAC TAT GTT TAC CTT GAT
        H   L   G   G   E   L   A   A   G   T   P   Y   Y   V   Y   L   D

 817  GAT GTA TCT CTC TAC GAT CCT AGG TTT GTA AAG CCT GTT GAA TAT GTA CTT
        D   V   S   L   Y   D   P   R   F   V   K   P   V   E   Y   V   L

 868  CCG CAG CCG GAT GTA CGT GTT AAC CAG GTA GGA TAC TTG CCG TTT GCA AAG
        P   Q   P   D   V   R   V   N   Q   V   G   Y   L   P   F   A   K

 919  AAG TAT GCT ACT GTT GTA TCT TCT TCA ACC AGC CCG CTT AAG TGG CAG CTT
        K   Y   A   T   V   V   S   S   S   T   S   P   L   K   W   Q   L

 970  CTC AAT TCG GCA AAT CAG GTT GTT TTG GAA GGT AAT ACA ATA CCA AAA GGA
        L   N   S   A   N   Q   V   V   L   E   G   N   T   I   P   K   G

1021  CTT GAC AAA GAT TCA CAG GAT TAT GTA CAT TGG ATA GAT TTC TCC AAC TTT
        L   D   K   D   S   Q   D   Y   V   H   W   I   D   F   S   N   F

1072  AAG ACT GAA GGA AAA GGT TAT TAC TTC AAG CTT CCG ACT GTA AAC AGC GAT
        K   T   E   G   K   G   Y   Y   F   K   L   P   T   V   N   S   D

1123  ACA AAT TAC AGC CAT CCT TTC GAT ATC AGT GCT GAT ATT TAC TCC AAG ATG
        T   N   Y   S   H   P   F   D   I   S   A   D   I   Y   S   K   M

1174  AAA TTT GAT GCA TTG GCA TTC TTC TAT CAC AAG AGA AGC GGT ATT CCT ATT
        K   F   D   A   L   A   F   F   Y   H   K   R   S   G   I   P   I

1225  GAA ATG CCG TAT GCA GGA GGA GAA CAG TGG ACC AGA CCT GCA GGA CAT ATT
        E   M   P   Y   A   G   G   E   Q   W   T   R   P   A   G   H   I

1276  GGA ATT GAG CCG AAC AAG GGA GAT ACA AAT GTT CCT ACA TGG CCT CAG GAT
        G   I   E   P   N   K   G   D   T   N   V   P   T   W   P   Q   D

1327  GAT GAA TAT GCA GGA AGA CCT CAA AAA TAT TAT ACA AAA GAT GTA ACC GGT
        D   E   Y   A   G   R   P   Q   K   Y   Y   T   K   D   V   T   G

1378  GGA TGG TAT GAT GCC GGT GAC CAC GGT AAA TAT GTT GTA AAC GGC GGT ATA
        G   W   Y   D   A   G   D   H   G   K   Y   V   V   N   G   G   I

1429  GCT GTT TGG ACA TTG ATG AAC ATG TAT GAA AGG GCA AAA ATC AGA GGC ATA
        A   V   W   T   L   M   N   M   Y   E   R   A   K   I   R   G   I

1480  GCT AAT CAA GGT GCT TAT AAA GAC GGT GGA ATG AAC ATA CCG GAG AGA AAT
        A   N   Q   G   A   Y   K   D   G   G   M   N   I   P   E   R   N

1531  AAC GGT TAT CCG GAC ATT CTT GAT GAA GCA AGA TGG GAA ATT GAG TTC TTT
        N   G   Y   P   D   I   L   D   E   A   R   W   E   I   E   F   F

1582  AAG AAA ATG CAG GTA ACT GAA AAA GAG GAT CCT TCC ATA GCC GGA ATG GTA
        K   K   M   Q   V   T   E   K   E   D   P   S   I   A   G   M   V

1633  CAC CAC AAA ATT CAC GAC TTC AGA TGG ACT GCT TTG GGT ATG TTG CCT CAC
        H   H   K   I   H   D   F   R   W   T   A   L   G   M   L   P   H

1684  GAA GAT CCC CAG CCA CGT TAC TTA AGG CCG GTA AGT ACG GCT GCG ACT TTG
        E   D   P   Q   P   R   Y   L   R   P   V   S   T   A   A   T   L
```

```
1735  AAC TTT GCG GCA ACT TTG GCA CAA AGT GCA CGT CTT TGG AAA GAT TAT GAT
        N   F   A   A   T   L   A   Q   S   A   R   L   W   K   D   Y   D

1786  CCG ACT TTT GCT GCT GAC TGT TTG GAA AAG GCT GAA ATA GCA TGG CAG GCG
        P   T   F   A   A   D   C   L   E   K   A   E   I   A   W   Q   A

1837  GCA TTA AAG CAT CCT GAT ATT TAT GCT GAG TAT ACT CCC GGT AGC GGT GGT
        A   L   K   H   P   D   I   Y   A   E   Y   T   P   G   S   G   G

1888  CCC GGA GGC GGA CCA TAC AAT GAC GAC TAT GTC GGA GAC GAA TTC TAC TGG
        P   G   G   G   P   Y   N   D   D   Y   V   G   D   E   F   Y   W

1939  GCA GCC TGC GAA CTT TAT GTA ACA ACA GGA AAA GAC GAA TAT AAG AAT TAC
        A   A   C   E   L   Y   V   T   T   G   K   D   E   Y   K   N   Y

1990  CTG ATG AAT TCA CCT CAC TAT CTT GAA ATG CCT GCA AAG ATG GGT GAA AAC
        L   M   N   S   P   H   Y   L   E   M   P   A   K   M   G   E   N

2041  GGT GGA GCA AAC GGA GAA GAC AAC GGA TTG TGG GGA TGC TTC ACC TGG GGA
        G   G   A   N   G   E   D   N   G   L   W   G   C   F   T   W   G

2092  ACT ACT CAA GGA TTG GGA ACC ATT ACT CTT GCA TTG GTT GAA AAC GGA TTG
        T   T   Q   G   L   G   T   I   T   L   A   L   V   E   N   G   L

2143  CCT GCT ACA GAC ATT CAA AAG GCA AGA AAC AAT ATA GCT AAA GCT GCT GAC
        P   A   T   D   I   Q   K   A   R   N   N   I   A   K   A   A   D

2194  AGA TGG CTT GAG AAT ATT GAA GAG CAA GGT TAC AGA CTG CCG ATC AAA CAG
        R   W   L   E   N   I   E   E   Q   G   Y   R   L   P   I   K   Q

2245  GCG GAG GAT GAG AGA GGC GGT TAT CCA TGG GGT TCA AAC TCC TTC ATT TTG
        A   E   D   E   R   G   G   Y   P   W   G   S   N   S   F   I   L

2296  AAC CAG ATG ATA GTT ATG GGA TAC GCA TAT GAC TTT ACA GGC AAC AGC AAG
        N   Q   M   I   V   M   G   Y   A   Y   D   F   T   G   N   S   K

2347  TAT CTT GAC GGA ATG CAG GAT GGT ATG AGC TAC CTG TTG GGA AGA AAC GGA
        Y   L   D   G   M   Q   D   G   M   S   Y   L   L   G   R   N   G

2398  CTG GAT CAG TCC TAT GTA ACA GGG TAT GGT GAG CGT CCA CTT CAG AAT CCT
        L   D   Q   S   Y   V   T   G   Y   G   E   R   P   L   Q   N   P

2449  CAT GAC AGA TTC TGG ACG CCG CAG ACA AGT AAG AAA TTC CCT GCT CCA CCT
        H   D   R   F   W   T   P   Q   T   S   K   K   F   P   A   P   P

2500  CCG GGT ATA ATT GCC GGT GGT CCG AAC TCC CGT TTC GAA GAC CCG ACA ATA
        P   G   I   I   A   G   G   P   N   S   R   F   E   D   P   T   I

2551  ACT GCA GCA GTT AAG AAG GAT ACA CCG CCG CAG AAG TGC TAC ATT GAC CAT
        T   A   A   V   K   K   D   T   P   P   Q   K   C   Y   I   D   H

2602  ACA GAC TCA TGG TCA ACC AAC GAG ATA ACT GTT AAC TGG AAT GCT CCG TTT
        T   D   S   W   S   T   N   E   I   T   V   N   W   N   A   P   F

2653  GCA TGG GTT ACA GCT TAT CTC GAT GAA ATT GAC TTA ATA ACA CCG CCA GGA
        A   W   V   T   A   Y   L   D   E   I   D   L   I   T   P   P   G

2704  GGA GTA GAC CCA GAA GAA CCG GAG GTT ATT TAT GGT GAC TGC AAT GGC GAC
        G   V   D   P   E   E   P   E   V   I   Y   G   D   C   N   G   D

2755  GGA AAA GTT AAT TCA ACT GAC GCT GTG GCA TTG AAG AGA TAT ATC TTG AGA
        G   K   V   N   S   T   D   A   V   A   L   K   R   Y   I   L   R
```

```
2806  TCA GGT ATA AGC ATC AAC ACT GAT AAT GCT GAT GTA AAT GCT GAT GGC AGA
        S   G   I   S   I   N   T   D   N   A   D   V   N   A   D   G   R

2857  GTT AAC TCT ACA GAC TTG GCA ATA TTG AAG AGA TAT ATT CTT AAA GAG ATA
        V   N   S   T   D   L   A   I   L   K   R   Y   I   L   K   E   I
                                •
2908  GAT GTA TTG CCA CAT AAA TAA AAA ACT GCC ATA TAA TGA TAA TCT GTT ATA
        D   V   L   P   H   K   -   K   T   A   I   -   -   -   S   V   I
```

Endoglucanase R (GH9)

```
   1  ACT TTA CCC ATT AAA AGG TTC GTG TAT TGC AAA ATA CAT ATC CGG TAT AAT
        T   L   P   I   K   R   F   V   Y   C   K   I   H   I   R   Y   N

  52  TAC ATT ATA TTA CCT GTA TAA CAT GAA GAC AAT GAT AAT ACA TAA AAA GAC
        Y   I   I   L   P   V   -   H   E   D   N   D   N   T   -   K   D

 103  AAC GTT TCA GAG AAT TAA ACA AAA TCA TTC CCT GTT ATT TGC CTC TTA TAT
        N   V   S   E   N   -   T   K   S   F   P   V   I   C   L   L   Y

 154  TTT TAA TAT TCA ATT GAG CAA AAA TTG CCG CAA GAA TGT TAT GTA TCA GCA
        F   -   Y   S   I   E   Q   K   L   P   Q   E   C   Y   V   S   A

 205  AAA TTA TAA TTT ATG ATA CAA TTT TTA TGA TAT AAT TAT GAT ATA ATT AAA
        K   L   -   F   M   I   Q   F   L   -   Y   N   Y   D   I   I   K

 256  AAT AGA ATC ATT TGT TCG CAT TTA TTT TAC TCG CAA ATA TCC AGT TAA TTC
        N   R   I   I   C   S   H   L   F   Y   S   Q   I   S   S   -   F

 307  ATG AAT TTC AAT ATC GCT ATT CTG AAG CTT ATC CAA AAC ACA AGT GTG AAC
        M   N   F   N   I   A   I   L   K   L   I   Q   N   T   S   V   N

 358  AAT CAA ATT TTA AAA AGG GGG AGA TAT ATA GTG AAA AAA CTC ATT ATC ACT
        N   Q   I   L   K   R   G   R   Y   I   V   K   K   L   I   I   T

 409  GTT ATA GTA TCT GCT GTC CTT TTA ACT GCT CTT ATA CCG CAG TTG CCT GTT
        V   I   V   S   A   V   L   L   T   A   L   I   P   Q   L   P   V

 460  TTT GCA GCA GAC TAT AAC TAT GGA GAA GCA CTC CAA AAA GCA ATT ATG TTC
        F   A   A   D   Y   N   Y   G   E   A   L   Q   K   A   I   M   F

 511  TAT GAA TTT CAA ATG TCC GGA AAG CTT CCC GAC AAC ATC CGT AAC AAC TGG
        Y   E   F   Q   M   S   G   K   L   P   D   N   I   R   N   N   W

 562  CGC GGT GAT TCA TGT CTC GGA GAC GGA AGC GAT GTA GGT CTT GAC CTC ACA
        R   G   D   S   C   L   G   D   G   S   D   V   G   L   D   L   T

 613  GGA GGT TGG TTT GAC GCC GGT GAC CAT GTA AAA TTC AAT CTG CCT ATG GCT
        G   G   W   F   D   A   G   D   H   V   K   F   N   L   P   M   A

 664  TAC ACA GCC ACT ATG CTT GCA TGG GCT GTG TAT GAG TAC AAG GAC GCG TTA
        Y   T   A   T   M   L   A   W   A   V   Y   E   Y   K   D   A   L

 715  CAA AAA AGC GGT CAA TTG GGC TAT TTA ATG GAT CAG ATT AAA TGG GCA TCG
        Q   K   S   G   Q   L   G   Y   L   M   D   Q   I   K   W   A   S

 766  GAC TAC TTC ATA AGA TGC CAT CCC GAA AAA TAT GTA TAT TAT TAT CAA GTG
        D   Y   F   I   R   C   H   P   E   K   Y   V   Y   Y   Y   Q   V

 817  GGT AAC GGT GAC ATG GAC CAC AGA TGG TGG GTG CCG GCA GAA TGT ATA GAT
```

```
        G   N   G   D   M   D   H   R   W   W   V   P   A   E   C   I   D

 868  GTT CAG GCA CCA AGA CCG TCT TAC AAA GTA GAT CTG TCA AAT CCC GGT TCC
        V   Q   A   P   R   P   S   Y   K   V   D   L   S   N   P   G   S

 919  ACA GTT ACT GCG GGT ACA GCT GCC GCA CTT GCT GCA ACT GCC TTG GTA TTC
        T   V   T   A   G   T   A   A   A   L   A   A   T   A   L   V   F

 970  AAA GGC ACT GAT CCG GCA TAT GCC GCT CTG TGC ATA CGT CAT GCA GAA GAA
        K   G   T   D   P   A   Y   A   A   L   C   I   R   H   A   E   E

1021  CTC TTT GAT TTT GCT GAA ACC ACT ATG AGT GAT AAA GGA TAT ACC GCA GCA
        L   F   D   F   A   E   T   T   M   S   D   K   G   Y   T   A   A

1072  TTG AAT TTC TAC ACA TCT CAC AGT GGA TGG TAT GAC GAG CTT TCC TGG GCA
        L   N   F   Y   T   S   H   S   G   W   Y   D   E   L   S   W   A

1123  GGT GCA TGG ATT TAT CTT GCA GAC GGT GAC GAA ACT TAT CTT GAA AAA GCT
        G   A   W   I   Y   L   A   D   G   D   E   T   Y   L   E   K   A

1174  GAA AAG TAT GTG GAT AAA TGG CCA ATC GAA AGC CAG ACA ACT TAC ATT GCT
        E   K   Y   V   D   K   W   P   I   E   S   Q   T   T   Y   I   A

1225  TAT TCA TGG GGT CAC TGC TGG GAC GAC GTT CAC TAC GGA GCA GCA CTT CTT
        Y   S   W   G   H   C   W   D   D   V   H   Y   G   A   A   L   L

1276  TTG GCA AAG ATT ACA AAC AAA TCC TTA TAC AAA GAA GCG ATA GAA AGA CAC
        L   A   K   I   T   N   K   S   L   Y   K   E   A   I   E   R   H

1327  CTG GAC TAT TGG ACA GTT GGA TTT AAT GGT CAG AGA GTC AGA TAT ACA CCA
        L   D   Y   W   T   V   G   F   N   G   Q   R   V   R   Y   T   P

1378  AAG GGT CTT GCT CAC CTC ACT GAC TGG GGT GTA TTA AGA CAT GCC ACT ACT
        K   G   L   A   H   L   T   D   W   G   V   L   R   H   A   T   T

1429  ACT GCA TTC CTT GCA TGT GTT TAT TCC GAC TGG TCA GAA TGT CCA AGG GAA
        T   A   F   L   A   C   V   Y   S   D   W   S   E   C   P   R   E

1480  AAA GCC AAT ATT TAC ATA GAT TTT GCC AAG AAA CAG GCT GAC TAT GCC TTA
        K   A   N   I   Y   I   D   F   A   K   K   Q   A   D   Y   A   L

1531  GGC AGC AGC GGC AGA AGT TAT GTA GTC GGA TTT GGT GTA AAT CCT CCG CAG
        G   S   S   G   R   S   Y   V   V   G   F   G   V   N   P   P   Q

1582  CAT CCG CAC CAC AGA ACT GCC CAC AGC TCA TGG TGT GAC AGT CAA AAA GTT
        H   P   H   H   R   T   A   H   S   S   W   C   D   S   Q   K   V

1633  CCT GAA TAC CAC AGA CAC GTT CTT TAC GGA GCA CTC GTA GGC GGA CCT GAT
        P   E   Y   H   R   H   V   L   Y   G   A   L   V   G   G   P   D

1684  GCC AGC GAT GCT TAT GTT GAT GAT ATA GGA AAC TAT GTA ACA AAT GAG GTT
        A   S   D   A   Y   V   D   D   I   G   N   Y   V   T   N   E   V

1735  GCC TGC GAC TAC AAT GCC GGT TTT GTA GGA TTG CTC GCC AAG ATG TAT GAA
        A   C   D   Y   N   A   G   F   V   G   L   L   A   K   M   Y   E

1786  AAA TAT GGC GGA AAC CCC ATA CCA AAC TTC ATG GCT ATA GAA GAA AAA ACA
        K   Y   G   G   N   P   I   P   N   F   M   A   I   E   E   K   T

1837  AAT GAA GAA ATT TAT GTT GAA GCT ACC GCC AAT TCA AAT AAC GGT GTC GAA
        N   E   E   I   Y   V   E   A   T   A   N   S   N   N   G   V   E

1888  TTG AAA ACA TAC CTT TAC AAT AAA TCC GGA TGG CCG GCA AGA GTT TGC GAC
```

```
          L   K   T   Y   L   Y   N   K   S   G   W   P   A   R   V   C   D

1939  AAG CTT TCC TTC AGA TAT TTC ATG GAC CTT ACG GAA TAT GTA TCC GCC GGA
          K   L   S   F   R   Y   F   M   D   L   T   E   Y   V   S   A   G

1990  TAC AAT CCT AAT GAT ATA ACT GTT TCT ATA ATT TAC AGT GCA GCA CCA ACT
          Y   N   P   N   D   I   T   V   S   I   I   Y   S   A   A   P   T

2041  GCA AAA ATT TCA AAA CCA ATA CTT TAT GAC GCA TCC AAA AAC ATA TAT TAT
          A   K   I   S   K   P   I   L   Y   D   A   S   K   N   I   Y   Y

2092  TGC GAA ATC GAT CTC TCC GGT ACC AAG ATA TTC CCC GGA AGC AAC TCA GAC
          C   E   I   D   L   S   G   T   K   I   F   P   G   S   N   S   D

2143  CAC CAG AAA GAA ACC CAA TTT AGA ATA CAG CCT CCT GCA GGC GCA CCT TGG
          H   Q   K   E   T   Q   F   R   I   Q   P   P   A   G   A   P   W

2194  GAC AAC ACC AAC GAC TTC TCC TAT CAG GGA ATC AAG AAA AAC GGT GAA GTT
          D   N   T   N   D   F   S   Y   Q   G   I   K   K   N   G   E   V

2245  GTA AAA GAA ATG CCT GTT TAT GAA GAC GGA GTT CTC ATA TTC GGT GTA GAA
          V   K   E   M   P   V   Y   E   D   G   V   L   I   F   G   V   E

2296  CCC AAT GGT ACC GGT CCT GCA ACA CCA ACG CCG AAA CCG TCC GTA AAT CCT
          P   N   G   T   G   P   A   T   P   T   P   K   P   S   V   N   P

2347  TCA CCT TCA CCT ACG CCA ACA TCG GAT ATT CTT TAC GGT GAC ATC AAT CTG
          S   P   S   P   T   P   T   S   D   I   L   Y   G   D   I   N   L

2398  GAC GGA AAA ATT AAC TCT TCA GAT GTT ACA CTG TTA AAA AGA TAT ATT GTG
          D   G   K   I   N   S   S   D   V   T   L   L   K   R   Y   I   V

2449  AAG TCC ATA GAT GTT TTC CCA ACC GCT GAT CCG GAA CGG AGC TTA ATA GCA
          K   S   I   D   V   F   P   T   A   D   P   E   R   S   L   I   A

2500  TCA GAT GTA AAC GGA GAC GGA AGG GTA AAC TCT ACA GAC TAT TCA TAC CTT
          S   D   V   N   G   D   G   R   V   N   S   T   D   Y   S   Y   L
                                                                      •
2551  AAA CGT TAT GTC TTG AAA ATC ATA CCA ACC ATA CCC GGA AAT TCA TGA GAT
          K   R   Y   V   L   K   I   I   P   T   I   P   G   N   S   -   D

2602  TAT CGT ATT CTT TAA AAC AAA AAG CTA AAG CTA AAA TAA TAA ACT AAA AAT
          Y   R   I   L   -   N   K   K   L   K   L   K   -   -   T   K   N

2653  TGA GAA GAT ATT TTA AGA TAA TCA AAA AAG AGA AGA GTT CAA AGC TCT TCT
          -   E   D   I   L   R   -   S   K   K   R   R   V   Q   S   S   S
```

Exoglucanase S (GH48)

```
  1  ATA TGT CAA ATT ATG TCA AAT GCG CGG CTG ATT TGA TAA AAA AGT TTG TTA
       I   C   Q   I   M   S   N   A   R   L   I   -   -   K   S   L   L

 52  ACA CAA ATT TAT TAT GTT AAC ACA AGT ATT TTT TGG GTC CAG CTT AGT TTT
       T   Q   I   Y   Y   V   N   T   S   I   F   W   V   Q   L   S   F

103  ATG ATG AAA ATA ATG CGT AAA ATT TAT CCG CCA AAA GGG GGA ATG AAT TTA
       M   M   K   I   M   R   K   I   Y   P   P   K   G   G   M   N   L

154  TTG CGG GTA GGT TGC ATT ATT TCA TCA TAT AAC TTA AAA AGA ATA AAA AAG
       L   R   V   G   C   I   I   S   S   Y   N   L   K   R   I   K   K

205  TAT ATT TGA AAG GGG AAG ATG GAG AGA ATG GTA AAA AGC AGA AAG ATT TCT
       Y   I   -   K   G   K   M   E   R   M   V   K   S   R   K   I   S

256  ATT CTG TTG GCA GTT GCA ATG CTG GTA TCC ATA ATG ATA CCC ACA ACT GCA
       I   L   L   A   V   A   M   L   V   S   I   M   I   P   T   T   A

307  TTC GCA GGT CCT ACA AAG GCA CCT ACA AAA GAT GGG ACA TCT TAT AAG GAT
       F   A   G   P   T   K   A   P   T   K   D   G   T   S   Y   K   D

358  CTT TTC CTT GAA CTC TAC GGA AAA ATT AAA GAT CCT AAG AAC GGA TAT TTC
       L   F   L   E   L   Y   G   K   I   K   D   P   K   N   G   Y   F

409  AGC CCA GAC GAG GGA ATT CCT TAT CAC TCA ATT GAA ACA TTG ATC GTT GAA
       S   P   D   E   G   I   P   Y   H   S   I   E   T   L   I   V   E

460  GCG CCG GAC TAC GGT CAC GTT ACT ACC AGT GAG GCT TTC AGC TAT TAT GTA
       A   P   D   Y   G   H   V   T   T   S   E   A   F   S   Y   Y   V

511  TGG CTT GAA GCA ATG TAT GGA AAT CTC ACA GGC AAC TGG TCC GGA GTA GAA
       W   L   E   A   M   Y   G   N   L   T   G   N   W   S   G   V   E

562  ACA GCA TGG AAA GTT ATG GAG GAT TGG ATA ATT CCT GAC AGC ACA GAG CAG
       T   A   W   K   V   M   E   D   W   I   I   P   D   S   T   E   Q

613  CCG GGT ATG TCT TCT TAC AAT CCA AAC AGC CCT GCC ACA TAT GCT GAC GAA
       P   G   M   S   S   Y   N   P   N   S   P   A   T   Y   A   D   E

664  TAT GAG GAT CCT TCA TAC TAT CCT TCA GAG TTG AAG TTT GAT ACC GTA AGA
       Y   E   D   P   S   Y   Y   P   S   E   L   K   F   D   T   V   R

715  GTT GGA TCC GAC CCT GTA CAC AAC GAC CTT GTA TCC GCA TAC GGT CCT AAC
       V   G   S   D   P   V   H   N   D   L   V   S   A   Y   G   P   N

766  ATG TAC CTC ATG CAC TGG TTG ATG GAC GTT GAC AAC TGG TAC GGT TTT GGT
       M   Y   L   M   H   W   L   M   D   V   D   N   W   Y   G   F   G

817  ACA GGA ACA CGG GCA ACA TTC ATA AAC ACC TTC CAA AGA GGT GAA CAG GAA
       T   G   T   R   A   T   F   I   N   T   F   Q   R   G   E   Q   E

868  TCC ACA TGG GAA ACC ATT CCT CAT CCG TCA ATA GAA GAG TTC AAA TAC GGC
       S   T   W   E   T   I   P   H   P   S   I   E   E   F   K   Y   G

919  GGA CCG AAC GGA TTC CTT GAT TTG TTT ACA AAG GAC AGA TCA TAT GCA AAA
       G   P   N   G   F   L   D   L   F   T   K   D   R   S   Y   A   K

970  CAG TGG CGT TAT ACA AAC GCT CCT GAC GCA GAA GGC CGT GCT ATA CAG GCT
       Q   W   R   Y   T   N   A   P   D   A   E   G   R   A   I   Q   A
```

```
1021  GTT TAC TGG GCA AAC AAA TGG GCA AAG GAG CAG GGT AAA GGT TCT GCC GTT
        V   Y   W   A   N   K   W   A   K   E   Q   G   K   G   S   A   V

1072  GCT TCC GTT GTA TCC AAG GCT GCA AAG ATG GGT GAC TTC TTG AGA AAC GAC
        A   S   V   V   S   K   A   A   K   M   G   D   F   L   R   N   D

1123  ATG TTC GAC AAA TAC TTC ATG AAG ATC GGT GCA CAG GAC AAG ACT CCT GCT
        M   F   D   K   Y   F   M   K   I   G   A   Q   D   K   T   P   A

1174  ACC GGT TAT GAC AGT GCA CAC TAC CTT ATG GCC TGG TAT ACT GCA TGG GGT
        T   G   Y   D   S   A   H   Y   L   M   A   W   Y   T   A   W   G

1225  GGT GGA ATT GGT GCA TCC TGG GCA TGG AAG ATC GGA TGC AGC CAC GCA CAC
        G   G   I   G   A   S   W   A   W   K   I   G   C   S   H   A   H

1276  TTC GGA TAT CAG AAC CCA TTC CAG GGA TGG GTA AGT GCA ACA CAG AGC GAC
        F   G   Y   Q   N   P   F   Q   G   W   V   S   A   T   Q   S   D

1327  TTT GCT CCT AAA TCA TCC AAC GGT AAG AGA GAC TGG ACA ACA AGC TAC AAG
        F   A   P   K   S   S   N   G   K   R   D   W   T   T   S   Y   K

1378  AGA CAG CTT GAA TTC TAT CAG TGG TTG CAG TCG GCT GAA GGT GGT ATT GCC
        R   Q   L   E   F   Y   Q   W   L   Q   S   A   E   G   G   I   A

1429  GGT GGA GCA ACC AAC TCC TGG AAC GGT AGA TAT GAG AAA TAT CCT GCT GGT
        G   G   A   T   N   S   W   N   G   R   Y   E   K   Y   P   A   G

1480  ACG TCA ACG TTC TAT GGT ATG GCA TAT GTT CCG CAT CCT GTA TAC GCT GAC
        T   S   T   F   Y   G   M   A   Y   V   P   H   P   V   Y   A   D

1531  CCG GGT AGT AAC CAG TGG TTC GGA TTC CAG GCA TGG TCA ATG CAG CGT GTA
        P   G   S   N   Q   W   F   G   F   Q   A   W   S   M   Q   R   V

1582  ATG GAG TAC TAC CTC GAA ACA GGA GAT TCA TCA GTT AAG AAT TTG ATT AAG
        M   E   Y   Y   L   E   T   G   D   S   S   V   K   N   L   I   K

1633  AAG TGG GTC GAC TGG GTA ATG AGC GAA ATT AAG CTC TAT GAC GAT GGA ACA
        K   W   V   D   W   V   M   S   E   I   K   L   Y   D   D   G   T

1684  TTT GCA ATT CCT AGC GAC CTC GAG TGG TCA GGT CAG CCT GAT ACA TGG ACC
        F   A   I   P   S   D   L   E   W   S   G   Q   P   D   T   W   T

1735  GGA ACA TAC ACA GGC AAC CCG AAC CTC CAT GTA AGA GTA ACT TCT TAC GGT
        G   T   Y   T   G   N   P   N   L   H   V   R   V   T   S   Y   G

1786  ACT GAC CTT GGT GTT GCA GGT TCA CTT GCA AAT GCT CTT GCA ACT TAT GCC
        T   D   L   G   V   A   G   S   L   A   N   A   L   A   T   Y   A

1837  GCA GCT ACA GAA AGA TGG GAA GGA AAA CTT GAT ACA AAA GCA AGA GAC ATG
        A   A   T   E   R   W   E   G   K   L   D   T   K   A   R   D   M

1888  GCT GCT GAA CTG GTT AAC CGT GCA TGG TAC AAC TTC TAC TGC TCT GAA GGA
        A   A   E   L   V   N   R   A   W   Y   N   F   Y   C   S   E   G

1939  AAA GGT GTT GTT ACT GAG GAA GCA CGT GCT GAC TAC AAA CGT TTC TTT GAG
        K   G   V   V   T   E   E   A   R   A   D   Y   K   R   F   F   E

1990  CAG GAA GTA TAC GTT CCG GCA GGT TGG AGC GGT ACT ATG CCG AAC GGT GAC
        Q   E   V   Y   V   P   A   G   W   S   G   T   M   P   N   G   D

2041  AAG ATT CAG CCT GGT ATT AAG TTC ATA GAC ATC CGT ACA AAA TAT AGA CAA
        K   I   Q   P   G   I   K   F   I   D   I   R   T   K   Y   R   Q
```

```
2092  GAT CCT TAC TAC GAT ATA GTA TAT CAG GCA TAC TTG AGA GGC GAA GCT CCT
        D   P   Y   Y   D   I   V   Y   Q   A   Y   L   R   G   E   A   P

2143  GTA TTG AAT TAT CAC CGC TTC TGG CAT GAA GTT GAC CTT GCA GTT GCA ATG
        V   L   N   Y   H   R   F   W   H   E   V   D   L   A   V   A   M

2194  GGT GTA TTG GCT ACA TAC TTC CCG GAT ATG ACA TAT AAA GTA CCT GGT ACT
        G   V   L   A   T   Y   F   P   D   M   T   Y   K   V   P   G   T

2245  CCT TCT ACT AAA TTA TAC GGC GAC GTC AAT GAT GAC GGA AAA GTT AAC TCA
        P   S   T   K   L   Y   G   D   V   N   D   D   G   K   V   N   S

2296  ACT GAC GCT GTA GCA TTG AAG AGA TAT GTT TTG AGA TCA GGT ATA AGC ATC
        T   D   A   V   A   L   K   R   Y   V   L   R   S   G   I   S   I

2347  AAC ACT GAC AAT GCC GAT TTG AAT GAA GAC GGC AGA GTT AAT TCA ACT GAC
        N   T   D   N   A   D   L   N   E   D   G   R   V   N   S   T   D

2398  TTA GGA ATT TTG AAG AGA TAT ATT CTC AAA GAA ATA GAT ACA TTG CCG TAC
        L   G   I   L   K   R   Y   I   L   K   E   I   D   T   L   P   Y
                •
2449  AAG AAC TAA TTT CAA AAC TGA TTT GAA AGG ACG GCT TGT GCC GGT CTT TTT
        K   N   -   F   Q   N   -   F   E   R   T   A   C   A   G   L   F

2500  TAC ATT TCT AAA GCC ATA CCA TGG CTT TTC GCA TAA TTT CTA TTA TAT TCG
        Y   I   S   K   A   I   P   W   L   F   A   -   F   L   L   Y   S
```

β-Glucosidase B (*Thermotoga neapolitana*)

```
   1  GTG GTC TCA GGC CTG TGG AGT TCT ATC TTA ACC TTG GAC TTC TTG CTT TTA
        V   V   S   G   L   W   S   S   I   L   T   L   D   F   L   L   L

  52  CAA CGT TCT TTC TGA TAT TCC TTG CGG TCA GGG AGA AAA AGA GGA GGG ATG
        Q   R   S   F   -   Y   S   L   R   S   G   R   K   R   G   G   M

 103  AAG ATG GAA AAG GTG AAT GAA ATC CTG TCT CAA CTC ACT CTG GAA GAA AAA
        K   M   E   K   V   N   E   I   L   S   Q   L   T   L   E   E   K

 154  AGT GAA ACT TGT AGT GGG GGA TGG ACT TCC GGG GTT GTT TGG AAA TCC CAT
        S   E   T   C   S   G   G   W   T   S   G   V   V   W   K   S   H

 205  TCC GGG TGG CGG TGC CGT GGG GAG ACG CAT CCT GTT CCA AGA GTG GGT CTT
        S   G   W   R   C   R   G   E   T   H   P   V   P   R   V   G   L

 256  CCT GCT TTC GTT CTG GCG GAT GGT CCA GCA GGA CTT AGA ATA AAT CCT ACA
        P   A   F   V   L   A   D   G   P   A   G   L   R   I   N   P   T

 307  AGA GAA AAC GAT GAG AAC ACC TAT TAC ACC ACC GCT TTT CCT GTT GAG ATC
        R   E   N   D   E   N   T   Y   Y   T   T   A   F   P   V   E   I

 358  ATG CTT GCC TCC ACC TGG AAC AGA GAG CTT CTC GAA GAA GTG GGA AAG GCA
        M   L   A   S   T   W   N   R   E   L   L   E   E   V   G   K   A

 409  ATG GGT GAA GAG GTG AGA GAG TAC GGT GTG GAT GTG CTC CTT GGT CCC GCG
        M   G   E   E   V   R   E   Y   G   V   D   V   L   L   G   P   A


 460  ATG AAC ATA CAC AGA AAT CCA CTT TGT GGA AGA AAC TTT GAA TAC TAC TCG
        M   N   I   H   R   N   P   L   C   G   R   N   F   E   Y   Y   S
```

```
 511  GAG GAT CCT GTC CTC TCC GGT GAA ATG GCC TCT TCC TTT GTG AAG GGA GTC
        E   D   P   V   L   S   G   E   M   A   S   S   F   V   K   G   V

 562  CAG TCA CAG GGA GTT GGT GCG TGT ATA AAG CAC TTC GTG GCG AAC AAC CAG
        Q   S   Q   G   V   G   A   C   I   K   H   F   V   A   N   N   Q

 613  GAG ACG AAC AGA ATG GTA GTG GAC ACG ATC GTG ATC GAA CGT GCT CTC AGA
        E   T   N   R   M   V   V   D   T   I   V   I   E   R   A   L   R

 664  GAG ATA TAT CTC AGG GGA TTC GAG ATC GCT GTG AAG AAA TCA AAA CCG TGG
        E   I   Y   L   R   G   F   E   I   A   V   K   K   S   K   P   W

 715  AGC GTG ATG AGT GCT TAC AAC AAA CTC AAT GGG AAG TAC TGC TCG CAG AAC
        S   V   M   S   A   Y   N   K   L   N   G   K   Y   C   S   Q   N

 766  GAG TGG CTC CTG AAG AAG GTT CTC AGG GAA GAG TGG GGT TTC GAA GGT TTC
        E   W   L   L   K   K   V   L   R   E   E   W   G   F   E   G   F

 817  GTT ATG AGT GAC TGG TAC GCT GGA GAC AAT CCC GTG GAA CAA CTC AAA GCA
        V   M   S   D   W   Y   A   G   D   N   P   V   E   Q   L   K   A

 868  GGC AAC GAT CTC ATC ATG CCT GGA AAG GCC TAC CAG GTG AAC ACA GAA CGA
        G   N   D   L   I   M   P   G   K   A   Y   Q   V   N   T   E   R

 919  AGA GAC GAA ATA GAA GAG ATC ATG GAG GCC CTG AAA GAA GGA AAA CTC AGC
        R   D   E   I   E   E   I   M   E   A   L   K   E   G   K   L   S

 970  GAA GAA GTT CTC GAT GAA TGT GTA AGA AAC ATC TTG AAA GTC CTT GTG AAC
        E   E   V   L   D   E   C   V   R   N   I   L   K   V   L   V   N

1021  GCA CCT TCT TTC AAA AAC TAC AGA TAC TCC AAC AAA CCC GAT CTT GAG AAG
        A   P   S   F   K   N   Y   R   Y   S   N   K   P   D   L   E   K

1072  CAC GCA AAG GTT GCT TAT GAA GCA GGA GCA GAA GGT GTT GTC CTT TTG AAA
        H   A   K   V   A   Y   E   A   G   A   E   G   V   V   L   L   K

1123  AAC GAA GAG GCT CTT CCT CTT TCT GAA AAC TCA AAG ATA GCC CTC TTT GGA
        N   E   E   A   L   P   L   S   E   N   S   K   I   A   L   F   G

1174  ACG GGC CAG ATC GAA ACG ATA AAA GGT GGA ACA GGA AGC GGC GAC ACC CAT
        T   G   Q   I   E   T   I   K   G   G   T   G   S   G   D   T   H

1225  CCA AGG TAC GCT ATT TCC ATC CTT GAG GGG ATA AAA GAA AGG GGT CTG AAT
        P   R   Y   A   I   S   I   L   E   G   I   K   E   R   G   L   N

1276  TTC GAC GAA GAA CTC GCA AAA ATC TAC GAA GAT TAC ATC AAG AAG ATG AGA
        F   D   E   E   L   A   K   I   Y   E   D   Y   I   K   K   M   R

1327  GAA ACA GAA GAG TAC AAA CCA AGA AGG GAT TCC TGG GGA ACG ATC ATA AAA
        E   T   E   E   Y   K   P   R   R   D   S   W   G   T   I   I   K

1378  CCA AAA CTT TCA GAA AAC TTC CTT TCG GAG AAG GAA GTA CAC AAA CTG GCA
        P   K   L   S   E   N   F   L   S   E   K   E   V   H   K   L   A

1429  AAA AAG AAC GAC GTG GCA GTC ATT GTG ATC AGC AGG ATT TCC GGA GAA GGC
        K   K   N   D   V   A   V   I   V   I   S   R   I   S   G   E   G

1480  TAT GAC AGA AAG CCG GTG AAG GGA GAC TTT TAC CTT TCT GAC GAT GAG ACT
        Y   D   R   K   P   V   K   G   D   F   Y   L   S   D   D   E   T

1531  GAT CTC ATA AAG ACT GTC TCC AGA GAG TTC CAT GAA CAG GGC AAG AAA GTG
        D   L   I   K   T   V   S   R   E   F   H   E   Q   G   K   K   V
```

```
1582  ATC GTT CTT CTC AAC ATA GGA AGT CCT GTT GAG GTT GTT AGC TGG AGA GAT
        I   V   L   L   N   I   G   S   P   V   E   V   V   S   W   R   D

1633  CTG GTG GAT GGG ATT CTC CTT GTG TGG CAG GCA GGG CAG GAA ACC GGC AGG
        L   V   D   G   I   L   L   V   W   Q   A   G   Q   E   T   G   R

1684  ATC GTT GCC GAT GTT CTC ACT GGA AGG ATC AAT CCA TCT GGA AAA CTT CCA
        I   V   A   D   V   L   T   G   R   I   N   P   S   G   K   L   P

1735  ACC ACC TTT CCG AGA GAC TAC TCT GAT GTA CCC TCC TGG ACC TTT CCT GGA
        T   T   F   P   R   D   Y   S   D   V   P   S   W   T   F   P   G

1786  GAG CCA AAG GAC AAT CCA CAG AAG GTG GTC TAC GAA GAG GAC ATC TAC GTG
        E   P   K   D   N   P   Q   K   V   V   Y   E   E   D   I   Y   V

1837  GGA TAC AGG TAC TAC GAC ACC TTC GGT GTG GAG CCG GCG TAC GAG TTC GGA
        G   Y   R   Y   Y   D   T   F   G   V   E   P   A   Y   E   F   G

1888  TAC GGC CTT TCT TAC ACG ACC TTT GAG TAC AGT GAC CTG AAC GTT TCG TTC
        Y   G   L   S   Y   T   T   F   E   Y   S   D   L   N   V   S   F

1939  GAC GGT GAA ACA CTC AGA GTT CAG TAC AGA ATA GAA AAC ACG GGC GGT CGT
        D   G   E   T   L   R   V   Q   Y   R   I   E   N   T   G   G   R

1990  GCA GGA AAG GAA GTC TCG CAG GTT TAC ATC AAG GCA CCG AAA GGA AAA ATC
        A   G   K   E   V   S   Q   V   Y   I   K   A   P   K   G   K   I

2041  GAC AAA CCC TTC CAG GAA CTC AAG GCG TTT CAC AAA ACC AGA CTT TTG AAT
        D   K   P   F   Q   E   L   K   A   F   H   K   T   R   L   L   N

2092  CCT GGA GAG TCT GAA GAA GTG GTG CTT GAG ATA CCT GTC AGA GAT CTT GCA
        P   G   E   S   E   E   V   V   L   E   I   P   V   R   D   L   A

2143  AGT TTC AAC GGT GAA GAA TGG GTT GTC GAA GCG GGT GAA TAC GAA GTA AGG
        S   F   N   G   E   E   W   V   V   E   A   G   E   Y   E   V   R

2194  GTT GGT GCG TCT TCG AGG AAC ATA AAA CTT AAA GGA ACG TTT TCC GTC GGT
        V   G   A   S   S   R   N   I   K   L   K   G   T   F   S   V   G
                                  •
2245  GAG GAG AGA AGA TTC AAA CCG TGA AGG AGG AGG GG
        E   E   R   R   F   K   P   -   R   R   R
```

Xyloglucanase A (GH74)

```
  1  AAA ACG TAA TTT GCA TCA ATG GTA TTT TAT TAC AAG TTT ATT AAT TAA CAA
       K   T   -   F   A   S   M   V   F   Y   Y   K   F   I   N   -   Q

 52  AAA TAA TTA ACC AAA TTA ATA AAA ATC AAT AAG TAA CGA AAG AAG AGT CAA
       K   -   L   T   K   L   I   K   I   N   K   -   R   K   K   S   Q

103  AAA AAG AAA GAT TGG TTT TTG CCG TCT TTC TTT TTT GAT TTT TAA AAT AAA
       K   K   K   D   W   F   L   P   S   F   F   F   D   F   -   N   K

154  AAA TTA ATA AAA TTA CGG GTA CAT CAA AGG AAA GTA CAG GTC CGA ATT TAT
       K   L   I   K   L   R   V   H   Q   R   K   V   Q   V   R   I   Y

205  ATA GCG AAG AGG GAC ATA ATT AAT CTT TTT AAT ATA GAC AAG GCA CCT GTT
       I   A   K   R   D   I   I   N   L   F   N   I   D   K   A   P   V

256  TGT TTA AAT ACC AAT GGG AGG GAG GTT GTT TTA TAA AAA GCG GTT GTA AAC
       C   L   N   T   N   G   R   E   V   V   L   -   K   A   V   V   N

307  AGG CAG GTT TGA GAG AAA ATT TTT AAT CAG TTT TTA AGG GAG GAG CGC TTA
       R   Q   V   -   E   K   I   F   N   Q   F   L   R   E   E   R   L

358  ATG GTT AAA AAG TTT ACA AGT AAA ATT AAG GCT GCT GTT TTT GCG GCT GTA
       M   V   K   K   F   T   S   K   I   K   A   A   V   F   A   A   V

409  GTT GCT GCA ACG GCA ATA TTT GGC CCC GCG ATT TCC AGC CAG GCT GTA ACC
       V   A   A   T   A   I   F   G   P   A   I   S   S   Q   A   V   T

460  AGC GTG CCT TAC AAA TGG GAC AAC GTG GTA ATC GGC GGA GGC GGA GGA TTT
       S   V   P   Y   K   W   D   N   V   V   I   G   G   G   G   G   F

511  ATG CCG GGT ATA GTT TTT AAT GAA ACG GAA AAG GAT TTG ATT TAT GCA CGT
       M   P   G   I   V   F   N   E   T   E   K   D   L   I   Y   A   R

562  GCC GAT ATC GGA GGA GCG TAC CGG TGG GAT CCT TCG ACC GAG ACA TGG ATT
       A   D   I   G   G   A   Y   R   W   D   P   S   T   E   T   W   I

613  CCG TTG CTG GAC CAT TTC CAA ATG GAT GAG TAC AGT TAT TAC GGA GTG GAA
       P   L   L   D   H   F   Q   M   D   E   Y   S   Y   Y   G   V   E

664  AGT ATT GCA ACC GAC CCT GTG GAT CCG AAC CGT GTT TAC ATA GTT GCA GGT
       S   I   A   T   D   P   V   D   P   N   R   V   Y   I   V   A   G

715  ATG TAT ACC AAC GAT TGG CTT CCT AAT ATG GGA GCA ATT CTT CGC TCA ACG
       M   Y   T   N   D   W   L   P   N   M   G   A   I   L   R   S   T

766  GAC AGG GGA GAA ACA TGG GAA AAA ACC ATA CTG CCT TTC AAG ATG GGC GGA
       D   R   G   E   T   W   E   K   T   I   L   P   F   K   M   G   G

817  AAC ATG CCG GGA AGA TCC ATG GGA GAA CGT CTT GCG ATC GAC CCG AAT GAC
       N   M   P   G   R   S   M   G   E   R   L   A   I   D   P   N   D

868  AAC AGG ATT CTT TAT CTT GGA ACA CGA TGC GGA AAC GGA CTT TGG AGA AGT
       N   R   I   L   Y   L   G   T   R   C   G   N   G   L   W   R   S

919  ACC GAC TAC GGT GTA ACA TGG TCC AAG GTT GAA AGT TTC CCA AAT CCC GGA
       T   D   Y   G   V   T   W   S   K   V   E   S   F   P   N   P   G

970  ACT TAC ATT TAT GAC CCG AAT TTT GAT TAT ACC AAA GAC ATT ATT GGA GTA
       T   Y   I   Y   D   P   N   F   D   Y   T   K   D   I   I   G   V
```

```
1021  GTC TGG GTT GTT TTT GAC AAG AGC AGC AGT ACA CCG GGC AAC CCT ACC AAG
        V   W   V   V   F   D   K   S   S   S   T   P   G   N   P   T   K

1072  ACT ATA TAT GTT GGT GTG GCT GAT AAA AAC GAA AGT ATT TAC CGC AGT ACG
        T   I   Y   V   G   V   A   D   K   N   E   S   I   Y   R   S   T

1123  GAC GGG GGT GTC ACC TGG AAA GCA GTT CCC GGA CAA CCT AAG GGA CTA CTT
        D   G   G   V   T   W   K   A   V   P   G   Q   P   K   G   L   L

1174  CCT CAC CAC GGG GTT TTG GCA TCC AAC GGA ATG TTG TAT ATA ACT TAT GGT
        P   H   H   G   V   L   A   S   N   G   M   L   Y   I   T   Y   G

1225  GAT ACC TGC GGT CCT TAT GAC GGC AAC GGA AAA GGT CAG GTT TGG AAG TTC
        D   T   C   G   P   Y   D   G   N   G   K   G   Q   V   W   K   F

1276  AAT ACA CGT ACA GGG GAA TGG ATA GAT ATC ACC CCG ATA CCT TAT TCA AGC
        N   T   R   T   G   E   W   I   D   I   T   P   I   P   Y   S   S

1327  AGT GAC AAT CGT TTC TGC TTT GCA GGA CTT GCA GTG GAC AGG CAG AAT CCT
        S   D   N   R   F   C   F   A   G   L   A   V   D   R   Q   N   P

1378  GAC ATT ATA ATG GTA ACT TCC ATG AAC GCG TGG TGG CCG GAT GAA TAT ATT
        D   I   I   M   V   T   S   M   N   A   W   W   P   D   E   Y   I

1429  TTC CGC AGT ACT GAC GGC GGA GCT ACA TGG AAG AAT ATC TGG GAA TGG GGA
        F   R   S   T   D   G   G   A   T   W   K   N   I   W   E   W   G

1480  ATG TAT CCT GAA CGT ATA CTG CAT TAT GAA ATA GAT ATT TCC GCA GCA CCG
        M   Y   P   E   R   I   L   H   Y   E   I   D   I   S   A   A   P

1531  TGG CTG GAT TGG GGA ACT GAG AAA CAG CTG CCG GAA ATC AAT CCG AAA CTG
        W   L   D   W   G   T   E   K   Q   L   P   E   I   N   P   K   L

1582  GGA TGG ATG ATA GGT GAC ATA GAG ATT GAC CCG TTT AAT TCC GAC CGC ATG
        G   W   M   I   G   D   I   E   I   D   P   F   N   S   D   R   M

1633  ATG TAT GTT ACC GGT GCA ACT ATC TAT GGT TGT GAC AAT CTT ACT GAC TGG
        M   Y   V   T   G   A   T   I   Y   G   C   D   N   L   T   D   W

1684  GAC AGA GGC GGC AAA GTA AAA ATC GAG GTA AAA GCT ACC GGA ATA GAA GAA
        D   R   G   G   K   V   K   I   E   V   K   A   T   G   I   E   E

1735  TGT GCG GTA TTA GAC CTG GTA AGC CCG CCG GAG GGT GCA CCG CTT GTA AGT
        C   A   V   L   D   L   V   S   P   P   E   G   A   P   L   V   S

1786  GCA GTT GGC GAC CTT GTC GGT TTT GTT CAT GAT GAC CTG AAA GTT GGT CCG
        A   V   G   D   L   V   G   F   V   H   D   D   L   K   V   G   P

1837  AAA AAA ATG CAC GTT CCT TCT TAT TCT TCA GGT ACG GGA ATT GAT TAT GCG
        K   K   M   H   V   P   S   Y   S   S   G   T   G   I   D   Y   A

1888  GAG CTT GTT CCG AAC TTT ATG GCA TTG GTT GCA AAG GCT GAT TTG TAT GAT
        E   L   V   P   N   F   M   A   L   V   A   K   A   D   L   Y   D

1939  GTA AAG AAG ATT TCT TTC TCT TAT GAC GGA GGA AGG AAT TGG TTC CAG CCA
        V   K   K   I   S   F   S   Y   D   G   G   R   N   W   F   Q   P

1990  CCT AAT GAA GCA CCA AAC TCG GTA GGC GGC GGT TCG GTT GCC GTT GCA GCC
        P   N   E   A   P   N   S   V   G   G   G   S   V   A   V   A   A

2041  GAT GCA AAA TCA GTT ATT TGG ACA CCG GAA AAT GCA AGT CCT GCA GTT ACA
        D   A   K   S   V   I   W   T   P   E   N   A   S   P   A   V   T
```

```
2092  ACG GAC AAC GGA AAC TCA TGG AAA GTT TGT ACA AAT CTT GGT ATG GGT GCG
        T   D   N   G   N   S   W   K   V   C   T   N   L   G   M   G   A

2143  GTG GTG GCA TCC GAC CGT GTG AAC GGT AAA AAA TTC TAC GCA TTC TAT AAC
        V   V   A   S   D   R   V   N   G   K   K   F   Y   A   F   Y   N

2194  GGC AAA TTC TAT ATA AGC ACG GAC GGT GGA TTA ACC TTT ACC GAT ACA AAG
        G   K   F   Y   I   S   T   D   G   G   L   T   F   T   D   T   K

2245  GCA CCG CAG CTT CCC AAG TCG GTT AAC AAG ATA AAA GCC GTA CCG GGC AAG
        A   P   Q   L   P   K   S   V   N   K   I   K   A   V   P   G   K

2296  GAA GGA CAT GTA TGG CTT GCT GCA AGA GAA GGC GGA TTG TGG AGG TCC ACT
        E   G   H   V   W   L   A   A   R   E   G   G   L   W   R   S   T

2347  GAC GGT GGA TAT ACG TTT GAG AAA CTC TCC AAT GTT GAC ACA GCT CAT GTG
        D   G   G   Y   T   F   E   K   L   S   N   V   D   T   A   H   V

2398  GTA GGC TTC GGA AAG GCA GCA CCG GGA CAG GAT TAC ATG GCG ATT TAC ATT
        V   G   F   G   K   A   A   P   G   Q   D   Y   M   A   I   Y   I

2449  ACC GGT AAA ATT GAC AAT GTT TTA GGA TTC TTC CGT TCC GAT GAT GCC GGC
        T   G   K   I   D   N   V   L   G   F   F   R   S   D   D   A   G

2500  AAG ACA TGG GTG CGT ATC AAC GAC GAC GAG CAC GGA TAT GGC GCT GTT GAT
        K   T   W   V   R   I   N   D   D   E   H   G   Y   G   A   V   D

2551  ACT GCA ATA ACA GGT GAC CCG AGA GTA TAC GGA CGT GTA TAT ATT GCC ACC
        T   A   I   T   G   D   P   R   V   Y   G   R   V   Y   I   A   T

2602  AAC GGA AGA GGT ATT GTT TAC GGC GAA CCT GCT TCA GAT GAG CCT GTA CCC
        N   G   R   G   I   V   Y   G   E   P   A   S   D   E   P   V   P

2653  ACT CCT CCG CAG GTT GAC AAA GGC CTG GTG GGC GAC TTG AAC GGT GAC AAT
        T   P   P   Q   V   D   K   G   L   V   G   D   L   N   G   D   N

2704  CGA ATA AAT TCA ACA GAC CTT ACT CTT ATG AAG AGA TAT ATC CTT AAA TCG
        R   I   N   S   T   D   L   T   L   M   K   R   Y   I   L   K   S

2755  ATA GAA GAT TTA CCT GTC GAA GAT GAT TTA TGG GCG GCG GAC ATA AAC GGC
        I   E   D   L   P   V   E   D   D   L   W   A   A   D   I   N   G

2806  GAC GGC AAA ATA AAT TCC ACA GAC TAT ACA TAC CTA AAG AAG TAT CTG CTT
        D   G   K   I   N   S   T   D   Y   T   Y   L   K   K   Y   L   L
                                          •
2857  CAA GCC ATT CCG GAG CTG CCG AAA AAA TAG TGA AAA CTT AGT CTC TCA CGA
        Q   A   I   P   E   L   P   K   K   -   -   K   L   S   L   S   R

2908  ACA ATG ATA AAT TTA AAA TAT CGT GTT AAC CCT GCA AAG TTT
        T   M   I   N   L   K   Y   R   V   N   P   A   K   F
```

Xylanase A (GH11)

```
  1  TAA AAA ATC CAT AAC ACT GTC AAT TTA CAG GCG AAG GGA AAG TAT ATA CTT
       -   K   I   H   N   T   V   N   L   Q   A   K   G   K   Y   I   L

 52  TTT CAG AAT TGC TTA AAA AAA TTA TTA ATA TAA ATT TAT TTT ATA ATT TAT
       F   Q   N   C   L   K   K   L   L   I   -   I   Y   F   I   I   Y

103  AAA AAA ATC TAA AAA AGG GGG AGA AAA ATG AAA CAA AAA TTA CTG GTA ACT
       K   K   I   -   K   R   G   R   K   M   K   Q   K   L   L   V   T

154  TTC CTG ATT TTA ATT ACT TTT ACC GTT TCA CTG ACT TTG TTT CCG GTA AAT
       F   L   I   L   I   T   F   T   V   S   L   T   L   F   P   V   N

205  GTA CGC GCT GAT GTA GTA ATT ACG TCA AAC CAG ACG GGT ACT CAT GGC GGG
       V   R   A   D   V   V   I   T   S   N   Q   T   G   T   H   G   G

256  TAC AAC TTT GAG TAC TGG AAA GAC ACC GGA AAC GGA ACC ATG GTC CTC AAA
       Y   N   F   E   Y   W   K   D   T   G   N   G   T   M   V   L   K

307  GAC GGT GGT GCG TTC AGC TGC GAA TGG AGC AAT ATC AAC AAT ATT CTT TTC
       D   G   G   A   F   S   C   E   W   S   N   I   N   N   I   L   F

358  CGT AAA GGT TTC AAA TAC GAT GAA ACA AAG ACA CAT GAT CAA CTT GGA TAC
       R   K   G   F   K   Y   D   E   T   K   T   H   D   Q   L   G   Y

409  ATA ACG GTA ACT TAT TCC TGC AAC TAT CAG CCA AAC GGA AAC TCT TAT CTG
       I   T   V   T   Y   S   C   N   Y   Q   P   N   G   N   S   Y   L

460  GGA GTC TAC GGA TGG ACC AGC AAT CCG CTT GTA GAG TAT TAC ATC ATC GAG
       G   V   Y   G   W   T   S   N   P   L   V   E   Y   Y   I   I   E

511  AGC TGG GGA ACC TGG AGA CCA CCG GGA GCA ACA CCA AAG GGC ACT ATT ACC
       S   W   G   T   W   R   P   P   G   A   T   P   K   G   T   I   T

562  GTT GAC GGT GGT ACA TAC GAG ATA TAC GAG ACC ACC AGA GTT AAC CAG CCT
       V   D   G   G   T   Y   E   I   Y   E   T   T   R   V   N   Q   P

613  TCC ATC AAA GGT ACA GCT ACT TTC CAG CAA TAC TGG AGT GTA CGT ACA TCA
       S   I   K   G   T   A   T   F   Q   Q   Y   W   S   V   R   T   S

664  AAA CGT ACA AGC GGA ACC ATA TCC GTA ACC GAA CAC TTT AAA GCC TGG GAA
       K   R   T   S   G   T   I   S   V   T   E   H   F   K   A   W   E

715  CGT CTG GGT ATG AAA ATG GGA AAA ATG TAT GAG GTT GCT TTG GTT GTA GAA
       R   L   G   M   K   M   G   K   M   Y   E   V   A   L   V   V   E

766  GGA TAC CAG AGC AGC GGA AAA GCC GAC GTA ACC AGC ATG ACA ATT ACT GTT
       G   Y   Q   S   S   G   K   A   D   V   T   S   M   T   I   T   V

817  GGC AAC GCA CCG TCA ACA TCA TCA CCA CCA GGT CCG ACA CCT GAA CCG ACT
       G   N   A   P   S   T   S   S   P   P   G   P   T   P   E   P   T

868  CCA AGA AGT GCT TTT TCA AAA ATC GAA GCT GAG GAG TAC AAC TCC CTC AAG
       P   R   S   A   F   S   K   I   E   A   E   E   Y   N   S   L   K

919  TCA TCA ACC ATT CAG ACC ATA GGC ACT TCC GAC GGA GGA AGC GGT ATA GGT
       S   S   T   I   Q   T   I   G   T   S   D   G   G   S   G   I   G

970  TAT ATT GAA AGC GGT GAC TAT CTG GTA TTT AAC AAA ATA AAC TTT GGA AAC
       Y   I   E   S   G   D   Y   L   V   F   N   K   I   N   F   G   N
```

```
1021  GGC GCA AAC TCT TTC AAG GCA AGG GTT GCA TCC GGT GCG GAC ACA CCC ACC
        G   A   N   S   F   K   A   R   V   A   S   G   A   D   T   P   T

1072  AAT ATC CAG TTA AGA CTC GGA AGC CCG ACC GGT ACT CTT ATA GGA ACT CTT
        N   I   Q   L   R   L   G   S   P   T   G   T   L   I   G   T   L

1123  ACG GTG GCT TCC ACA GGT GGT TGG AAC AAT TAC GAG GAA AAA TCC TGC AGC
        T   V   A   S   T   G   G   W   N   N   Y   E   E   K   S   C   S

1174  ATA ACC AAC ACT ACA GGA CAG CAC GAC TTA TAT CTG GTA TTC TCA GGT CCT
        I   T   N   T   T   G   Q   H   D   L   Y   L   V   F   S   G   P

1225  GTT AAC ATT GAC TAC TTC ATA TTC GAC TCG AAT GGC GTA AAT CCT ACA CCC
        V   N   I   D   Y   F   I   F   D   S   N   G   V   N   P   T   P

1276  ACC TCT CAG CCT CAA CAA GGC CAG GTT TTG GGT GAC TTG AAC GGA GAC AAA
        T   S   Q   P   Q   Q   G   Q   V   L   G   D   L   N   G   D   K

1327  CAA GTA AAT TCA ACA GAC TAC ACA GCA CTG AAG AGA CAT TTG CTC AAT ATA
        Q   V   N   S   T   D   Y   T   A   L   K   R   H   L   L   N   I

1378  ACC AGA CTT TCA GGA ACT GCT CTT GCC AAC GCC GAT TTA AAC GGT GAC GGC
        T   R   L   S   G   T   A   L   A   N   A   D   L   N   G   D   G

1429  AAA GTT GAT TCC ACT GAC CTT ATG ATT CTA CAC AGA TAT CTT CTC GGT ATA
        K   V   D   S   T   D   L   M   I   L   H   R   Y   L   L   G   I

1480  ATT TCA TCT TTT CCA CGC AGC AAT CCA CAA CCA AGC AGT AAC CCT CAA CCA
        I   S   S   F   P   R   S   N   P   Q   P   S   S   N   P   Q   P

1531  AGC AGC AAT CCG CAG CCA ACG ATT AAT CCA AAT GCG AAA CTG GTG GCT CTT
        S   S   N   P   Q   P   T   I   N   P   N   A   K   L   V   A   L

1582  ACC TTT GAC GAC GGT CCG GAC AAC GTA CTT ACG GCA CGG GTT CTC GAC AAG
        T   F   D   D   G   P   D   N   V   L   T   A   R   V   L   D   K

1633  CTT GAT AAA TAT AAC GTT AAG GCT ACA TTC ATG GTA GTA GGT CAG AGA GTC
        L   D   K   Y   N   V   K   A   T   F   M   V   V   G   Q   R   V

1684  AAT GAT TCG ACG GCT GCC ATC ATC AGA AGG ATG GTT AAT TCA GGC CAT GAA
        N   D   S   T   A   A   I   I   R   R   M   V   N   S   G   H   E

1735  ATA GGA AAC CAC TCA TGG AGT TAT TCA GGC ATG GCC AAT ATG AGT CCG GAT
        I   G   N   H   S   W   S   Y   S   G   M   A   N   M   S   P   D

1786  CAG ATA AGG AAA TCC ATT GCC GAT ACA AAT GCA GTT ATT CAA AAA TAT GCT
        Q   I   R   K   S   I   A   D   T   N   A   V   I   Q   K   Y   A

1837  GGA ACA ACA CCC AAG TTC TTC CGT CCG CCG AAC CTC GAA ACA AGC CCA ACA
        G   T   T   P   K   F   F   R   P   P   N   L   E   T   S   P   T

1888  TTA TTC AAC AAT GTT GAC TTG GTG TTT GTC GGC GGC TTA ACG GCA AAT GAC
        L   F   N   N   V   D   L   V   F   V   G   G   L   T   A   N   D

1939  TGG ATT CCA TCC ACA ACC GCC GAA CAG AGG GCT GCC GCA GTT ATA AAC GGT
        W   I   P   S   T   T   A   E   Q   R   A   A   A   V   I   N   G

1990  GTC AGA GAC GGT ACA ATA ATT CTT TTG CAT GAT GTT CAA CCT GAG CCA CAC
        V   R   D   G   T   I   I   L   L   H   D   V   Q   P   E   P   H

2041  CCG ACA CCG GAA GCT CTG GAT ATA ATC ATC CCT ACA CTT AAG AGC CGG GGC
        P   T   P   E   A   L   D   I   I   I   P   T   L   K   S   R   G
```

```
2092  TAT GAA TTT GTG ACC TTG ACT GAG TTG TTC ACG TTA AAG GGT GTG CCA ATT
        Y   E   F   V   T   L   T   E   L   F   T   L   K   G   V   P   I
                                                        ●
2143  GAC CCA TCA GTC AAA AGA ATG TAT AAC TCT GTA CCG TAA GAC AAA TCC GAA
        D   P   S   V   K   R   M   Y   N   S   V   P   -   D   K   S   E

2194  TAT CAA ATC AGT ACA ACA AAG GGG GAT AAA ATT ACT CCC CCT TTT TTT GCT
        Y   Q   I   S   T   T   K   G   D   K   I   T   P   P   F   F   A

2245  TTA TTG CCC CGT TAT TCT CTC TTT TTA TAT ATT TTT ATC GAT ATT GCG ATG
        L   L   P   R   Y   S   L   F   L   Y   I   F   I   D   I   A   M

2296  GAT ATT AAA TTT ACA ACC AAC GTA AAT ATT ACA GTT GCA ACG AAT GCC GTA
        D   I   K   F   T   T   N   V   N   I   T   V   A   T   N   A   V

2347  ACA ACG TTT ACA TAT GAT GCC GAA
        T   T   F   T   Y   D   A   E
```

Xylanase C (GH10)

```
   1  ACA AAC AAA TCA AAA TAA GTT ATA ATT TAA TTT AGG ATC ATT CAG AGA AAA
        T   N   K   S   K   -   V   I   I   -   F   R   I   I   Q   R   K

  52  ATT TAA CTT TAT TGT AAC AAG GCA ATA AAT GGC CCC ACT AAT TTC TTA TAA
        I   -   L   Y   C   N   K   A   I   N   G   P   T   N   F   L   -

 103  AGT ACA GGG TGC GGT CTG TAT TTT CGA CGG TGG GTG CGT TTT TGT ATG GTA
        S   T   G   C   G   L   Y   F   R   R   W   V   R   F   C   M   V

 154  AAA AAC TGT TGA CAT GGG AAA ACA CCC AAG CTT TAT GAA AAG CTT GGA TAT
        K   N   C   -   H   G   K   T   P   K   L   Y   E   K   L   G   Y

 205  ATA AAA TAA TTT TAA ATT TTG GGA GGT AGA TCT ATG CTG AAG AAA AAA CTG
        I   K   -   F   -   I   L   G   G   R   S   M   L   K   K   K   L

 256  TTG ACC CTT TTG ACA GTC TTT GCT CTG CTG ACT GTC GGT ATC TGC GGA AGT
        L   T   L   L   T   V   F   A   L   L   T   V   G   I   C   G   S

 307  TTT TTG CCG TTA CCC AAA GCA TCC GCA GCA GCT CTG ATT TAC GAT GAT TTT
        F   L   P   L   P   K   A   S   A   A   A   L   I   Y   D   D   F

 358  GAA ACA GGT CTG AAC GGA TGG GGA CCA AGA GGA CCG GAA ACC GTC GAA CTT
        E   T   G   L   N   G   W   G   P   R   G   P   E   T   V   E   L

 409  ACC ACC GAG GAA GCT TAC TCG GGA AGA TAC AGT TTG AAG GTC AGC GGA CGT
        T   T   E   E   A   Y   S   G   R   Y   S   L   K   V   S   G   R

 460  ACC AGC ACA TGG AAC GGG CCC ATG GTT GAC AAA ACC GAT GTG TTG ACT TTG
        T   S   T   W   N   G   P   M   V   D   K   T   D   V   L   T   L

 511  GGC GAA AGC TAT AAG TTG GGC GTA TAT GTA AAA TTC GTG GGT GAT TCC TAT
        G   E   S   Y   K   L   G   V   Y   V   K   F   V   G   D   S   Y

 562  TCA AAT GAG CAA AGA TTC AGT TTG CAG CTT CAA TAT AAC GAC GGA GCA GGA
        S   N   E   Q   R   F   S   L   Q   L   Q   Y   N   D   G   A   G

 613  GAT GTA TAC CAA AAT ATA AAA ACC GCC ACG GTT TAC AAG GGA ACA TGG ACT
        D   V   Y   Q   N   I   K   T   A   T   V   Y   K   G   T   W   T
 664  TTG CTG GAA GGA CAG CTT ACA GTT CCC AGC CAT GCA AAG GAC GTA AAA ATA
        L   L   E   G   Q   L   T   V   P   S   H   A   K   D   V   K   I
```

```
 715  TAT GTG GAA ACC GAA TTT AAA AAT TCT CCG AGT CCG CAG GAC TTG ATG GAT
        Y   V   E   T   E   F   K   N   S   P   S   P   Q   D   L   M   D

 766  TTC TAT ATT GAC GAT TTC ACA GCA ACA CCT GCA AAT TTG CCT GAA ATT GAG
        F   Y   I   D   D   F   T   A   T   P   A   N   L   P   E   I   E

 817  AAA GAT ATT CCA AGC TTG AAA GAT GTC TTT GCC GGT TAT TTC AAA GTG GGT
        K   D   I   P   S   L   K   D   V   F   A   G   Y   F   K   V   G

 868  GGT GCC GCA ACT GTG GCG GAA CTG GCG CCG AAG CCT GCA AAA GAG CTT TTC
        G   A   A   T   V   A   E   L   A   P   K   P   A   K   E   L   F

 919  CTC AAG CAT TAT AAC AGC TTG ACT TTT GGT AAT GAG TTA AAA CCG GAA AGT
        L   K   H   Y   N   S   L   T   F   G   N   E   L   K   P   E   S

 970  GTA CTT GAC TAT GAT GCT ACA ATT GCT TAT ATG GAG GCA AAC GGA GGC GAC
        V   L   D   Y   D   A   T   I   A   Y   M   E   A   N   G   G   D

1021  CAG GTT AAT CCG CAG ATA ACC TTG AGA GCG GCA AGA CCC CTG TTG GAG TTT
        Q   V   N   P   Q   I   T   L   R   A   A   R   P   L   L   E   F

1072  GCG AAA GAA CAC AAC ATA CCT GTA AGA GGA CAT ACC CTT GTA TGG CAC AGC
        A   K   E   H   N   I   P   V   R   G   H   T   L   V   W   H   S

1123  CAG ACA CCG GAC TGG TTC TTC AGA GAA AAT TAC TCT CAG GAC GAA AAT GCT
        Q   T   P   D   W   F   F   R   E   N   Y   S   Q   D   E   N   A

1174  CCC TGG GCA TCC AAG GAA GTA ATG CTG CAA AGG TTG GAA AAC TAC ATA AAG
        P   W   A   S   K   E   V   M   L   Q   R   L   E   N   Y   I   K

1225  AAT TTA ATG GAA GCT TTG GCG ACC GAA TAT CCG ACG GTT AAG TTC TAT GCA
        N   L   M   E   A   L   A   T   E   Y   P   T   V   K   F   Y   A

1276  TGG GAC GTT GTG AAT GAG GCT GTT GAT CCT AAT ACT TCA GAC GGT ATG AGA
        W   D   V   V   N   E   A   V   D   P   N   T   S   D   G   M   R

1327  ACT CCG GGT TCG AAT AAC AAA AAT CCC GGA AGC TCC CTG TGG ATG CAA ACC
        T   P   G   S   N   N   K   N   P   G   S   S   L   W   M   Q   T

1378  GTT GGA AGA GAT TTT ATT GTT AAA GCT TTT GAA TAT GCA AGA AAA TAT GCT
        V   G   R   D   F   I   V   K   A   F   E   Y   A   R   K   Y   A

1429  CCT GCG GAT TGT AAA CTC TTC TAC AAT GAC TAT AAT GAA TAT GAA GAC AGA
        P   A   D   C   K   L   F   Y   N   D   Y   N   E   Y   E   D   R

1480  AAA TGT GAT TTT ATT ATT GAA ATT CTT ACC GAA CTT AAA GCC AAA GGC CTG
        K   C   D   F   I   I   E   I   L   T   E   L   K   A   K   G   L

1531  GTT GAC GGT ATG GGT ATG CAA TCC CAC TGG GTT ATG GAT TAT CCA AGC ATA
        V   D   G   M   G   M   Q   S   H   W   V   M   D   Y   P   S   I

1582  AGC ATG TTT GAA AAA TCC ATC AGA AGA TAT GCA GCA TTG GGA TTG GAA ATT
        S   M   F   E   K   S   I   R   R   Y   A   A   L   G   L   E   I

1633  CAG CTT ACC GAG CTG GAT ATA AGA AAT CCT GAC AAC AGC CAG TGG GCT TTG
        Q   L   T   E   L   D   I   R   N   P   D   N   S   Q   W   A   L

1684  GAA CGT CAG GCT AAT CGT TAT AAG GAG CTT GTA ACA AAA TTG GTC GAT TTG
        E   R   Q   A   N   R   Y   K   E   L   V   T   K   L   V   D   L

1735  AAA AAA GAA GGC ATA AAC ATT ACG GCA TTG GTA TTC TGG GGA ATA ACC GAC
        K   K   E   G   I   N   I   T   A   L   V   F   W   G   I   T   D
```

```
1786  GCG ACA AGC TGG CTT GGA GGA TAT CCG CTC CTG TTT GAC GCG GAA TAC AAG
        A   T   S   W   L   G   G   Y   P   L   L   F   D   A   E   Y   K

1837  GCA AAA CCT GCA TTT TAT GCT ATA GTT AAC AGC GTT CCG CCG CTT CCG ACA
        A   K   P   A   F   Y   A   I   V   N   S   V   P   P   L   P   T

1888  GAA CCG CCG GTT CAG GTT ATA CCC GGT GAC GTA AAC GGT GAC GGT CGT GTA
        E   P   P   V   Q   V   I   P   G   D   V   N   G   D   G   R   V

1939  AAT TCA TCC GAC TTG ACT CTT ATG AAA AGA TAC CTT TTA AAA TCC ATA AGC
        N   S   S   D   L   T   L   M   K   R   Y   L   L   K   S   I   S

1990  GAC TTC CCG ACA CCG GAA GGA AAA ATT GCG GCG GAT TTA AAC GAA GAC GGC
        D   F   P   T   P   E   G   K   I   A   A   D   L   N   E   D   G

2041  AAG GTA AAC TCG ACA GAT TTG TTA GCG CTG AAA AAA CTC GTT CTG AGA GAA
        K   V   N   S   T   D   L   L   A   L   K   K   L   V   L   R   E
                •
2092  CTT TGA TCA AAA ACA GAT GAA AAC CGT TGT GCT TGC ATG TCT GCC TAA TAG
        L   -   S   K   T   D   E   N   R   C   A   C   M   S   A   -   -
```

Xylanase Z (GH10)

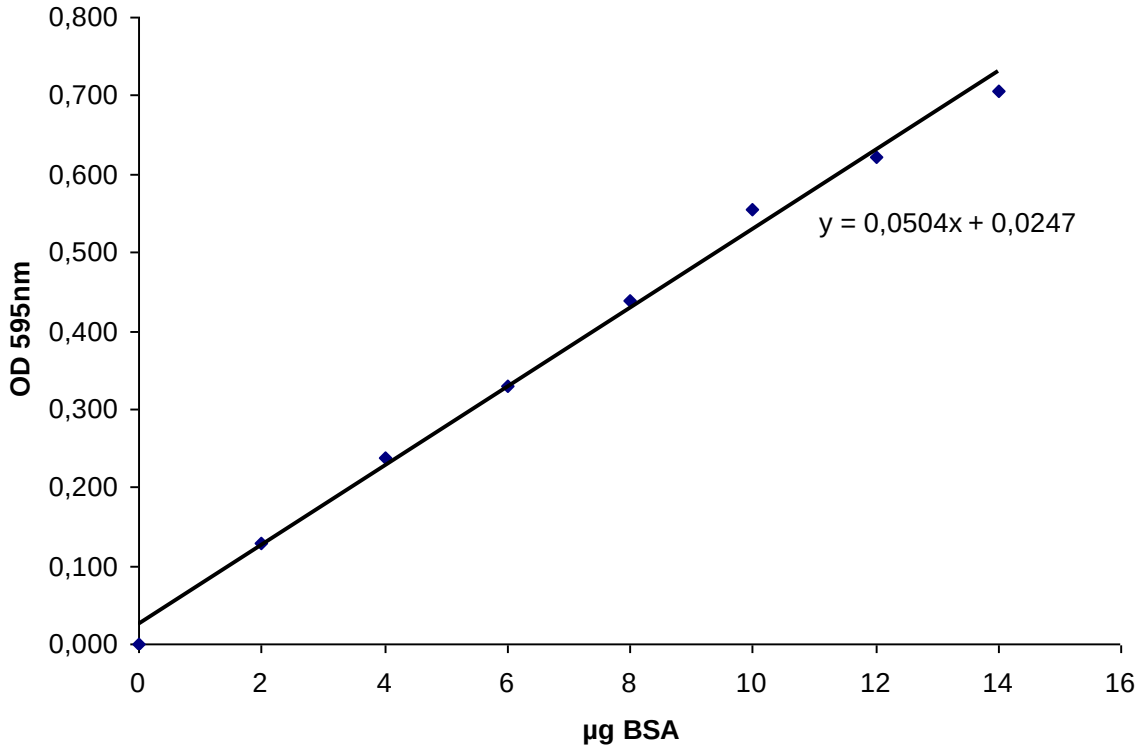
```
   1  TAT ATA AAT AAG GGT ATT AAT TCT GCA AAA AGA AAA GTG TTT GCT ACA TGA
        Y   I   N   K   G   I   N   S   A   K   R   K   V   F   A   T   -

  52  GGT CCA TTA ATT TTT ATT TTA TAT CAT AAA TCA AAA AGG AGG AGA AAC ATG
        G   P   L   I   F   I   L   Y   H   K   S   K   R   R   R   N   M

 103  TCA AGA AAA CTT TTC AGT GTA TTA CTT GTT GGC TTG ATG CTT ATG ACA TCG
        S   R   K   L   F   S   V   L   L   V   G   L   M   L   M   T   S

 154  TTG CTT GTC ACA ATA AGC AGT ACA TCA GCG GCA TCC TTG CCA ACC ATG CCG
        L   L   V   T   I   S   S   T   S   A   A   S   L   P   T   M   P

 205  CCT TCG GGA TAT GAC CAG GTA AGG AAC GGC GTT CCG AGA GGG CAG GTC GTA
        P   S   G   Y   D   Q   V   R   N   G   V   P   R   G   Q   V   V

 256  AAT ATT TCT TAT TTC TCC ACG GCC ACC AAC AGT ACC AGG CCG GCA AGA GTT
        N   I   S   Y   F   S   T   A   T   N   S   T   R   P   A   R   V

 307  TAT TTG CCG CCG GGA TAT TCA AAG GAC AAA AAA TAC AGT GTT TTG TAT CTC
        Y   L   P   P   G   Y   S   K   D   K   K   Y   S   V   L   Y   L

 358  TTA CAC GGC ATA GGC GGT AGT GAA AAC GAC TGG TTC GAA GGG GGA GGC AGA
        L   H   G   I   G   G   S   E   N   D   W   F   E   G   G   G   R

 409  GCC AAT GTT ATT GCC GAC AAT CTG ATT GCC GAG GGA AAA ATC AAG CCC CTG
        A   N   V   I   A   D   N   L   I   A   E   G   K   I   K   P   L

 460  ATA ATT GTA ACA CCG AAT ACT AAC GCC GCC GGT CCG GGA ATA GCG GAC GGT
        I   I   V   T   P   N   T   N   A   A   G   P   G   I   A   D   G

 511  TAT GAA AAT TTC ACA AAA GAT TTG CTC AAC AGT CTT ATT CCC TAT ATC GAA
        Y   E   N   F   T   K   D   L   L   N   S   L   I   P   Y   I   E

 562  TCT AAC TAT TCA GTC TAC ACC GAC CGC GAA CAT CGG GCG ATT GCA GGA CTT
        S   N   Y   S   V   Y   T   D   R   E   H   R   A   I   A   G   L
```

```
 613  TCA ATG GGT GGA GGA CAA TCG TTT AAT ATT GGA TTG ACC AAT CTC GAT AAA
        S   M   G   G   G   Q   S   F   N   I   G   L   T   N   L   D   K

 664  TTT GCC TAT ATT GGC CCG ATT TCA GCG GCT CCA AAC ACT TAT CCA AAT GAG
        F   A   Y   I   G   P   I   S   A   A   P   N   T   Y   P   N   E

 715  AGG CTT TTT CCT GAC GGA GGA AAA GCT GCA AGG GAG AAA TTG AAA CTG CTC
        R   L   F   P   D   G   G   K   A   A   R   E   K   L   K   L   L

 766  TTT ATT GCC TGC GGA ACC AAT GAC AGT CTG ATA GGT TTT GGA CAG AGA GTA
        F   I   A   C   G   T   N   D   S   L   I   G   F   G   Q   R   V

 817  CAT GAA TAT TGC GTT GCC AAC AAC ATT AAC CAT GTC TAT TGG CTT ATT CAG
        H   E   Y   C   V   A   N   N   I   N   H   V   Y   W   L   I   Q

 868  GGC GGA GGA CAC GAT TTT AAT GTG TGG AAG CCC GGA TTG TGG AAT TTC CTT
        G   G   G   H   D   F   N   V   W   K   P   G   L   W   N   F   L

 919  CAA ATG GCA GAT GAA GCC GGA TTG ACG AGG GAT GGA AAC ACT CCG GTT CCG
        Q   M   A   D   E   A   G   L   T   R   D   G   N   T   P   V   P

 970  ACA CCC AGT CCA AAG CCG GCT AAC ACA CGT ATT GAA GCG GAA GAT TAT GAC
        T   P   S   P   K   P   A   N   T   R   I   E   A   E   D   Y   D

1021  GGT ATT AAT TCT TCA AGT ATT GAG ATA ATA GGT GTT CCA CCT GAA GGA GGC
        G   I   N   S   S   S   I   E   I   I   G   V   P   P   E   G   G

1072  AGA GGA ATA GGT TAT ATT ACC AGT GGT GAT TAT CTG GTA TAC AAG AGT ATA
        R   G   I   G   Y   I   T   S   G   D   Y   L   V   Y   K   S   I

1123  GAC TTT GGA AAC GGA GCA ACG TCG TTT AAG GCC AAG GTT GCA AAT GCA AAT
        D   F   G   N   G   A   T   S   F   K   A   K   V   A   N   A   N

1174  ACT TCC AAT ATT GAA CTT AGA TTA AAC GGT CCG AAT GGT ACT CTC ATA GGC
        T   S   N   I   E   L   R   L   N   G   P   N   G   T   L   I   G

1225  ACA CTC TCG GTA AAA TCC ACA GGA GAT TGG AAT ACA TAT GAG GAG CAA ACT
        T   L   S   V   K   S   T   G   D   W   N   T   Y   E   E   Q   T

1276  TGC AGC ATT AGC AAA GTC ACC GGA ATA AAT GAT TTG TAC TTG GTA TTC AAA
        C   S   I   S   K   V   T   G   I   N   D   L   Y   L   V   F   K

1327  GGC CCT GTA AAC ATA GAC TGG TTC ACT TTT GGC GTT GAA AGC AGT TCC ACA
        G   P   V   N   I   D   W   F   T   F   G   V   E   S   S   S   T

1378  GGT CTG GGG GAT TTA AAT GGT GAC GGA AAT ATT AAC TCG TCG GAC CTT CAG
        G   L   G   D   L   N   G   D   G   N   I   N   S   S   D   L   Q

1429  GCG TTA AAG AGG CAT TTG CTC GGT ATA TCA CCG CTT ACG GGA GAG GCT CTT
        A   L   K   R   H   L   L   G   I   S   P   L   T   G   E   A   L

1480  TTA AGA GCG GAT GTA AAT AGG AGC GGC AAA GTG GAT TCT ACT GAC TAT TCA
        L   R   A   D   V   N   R   S   G   K   V   D   S   T   D   Y   S

1531  GTG CTG AAA AGA TAT ATA CTC CGC ATT ATT ACA GAG TTC CCC GGA CAA GGT
        V   L   K   R   Y   I   L   R   I   I   T   E   F   P   G   Q   G

1582  GAT GTA CAG ACA CCC AAT CCG TCT GTT ACT CCG ACA CAA ACT CCT ATC CCC
        D   V   Q   T   P   N   P   S   V   T   P   T   Q   T   P   I   P

1633  ACG ATT TCG GGA AAT GCT CTT AGG GAT TAT GCG GAG GCA AGG GGA ATA AAA
        T   I   S   G   N   A   L   R   D   Y   A   E   A   R   G   I   K
```

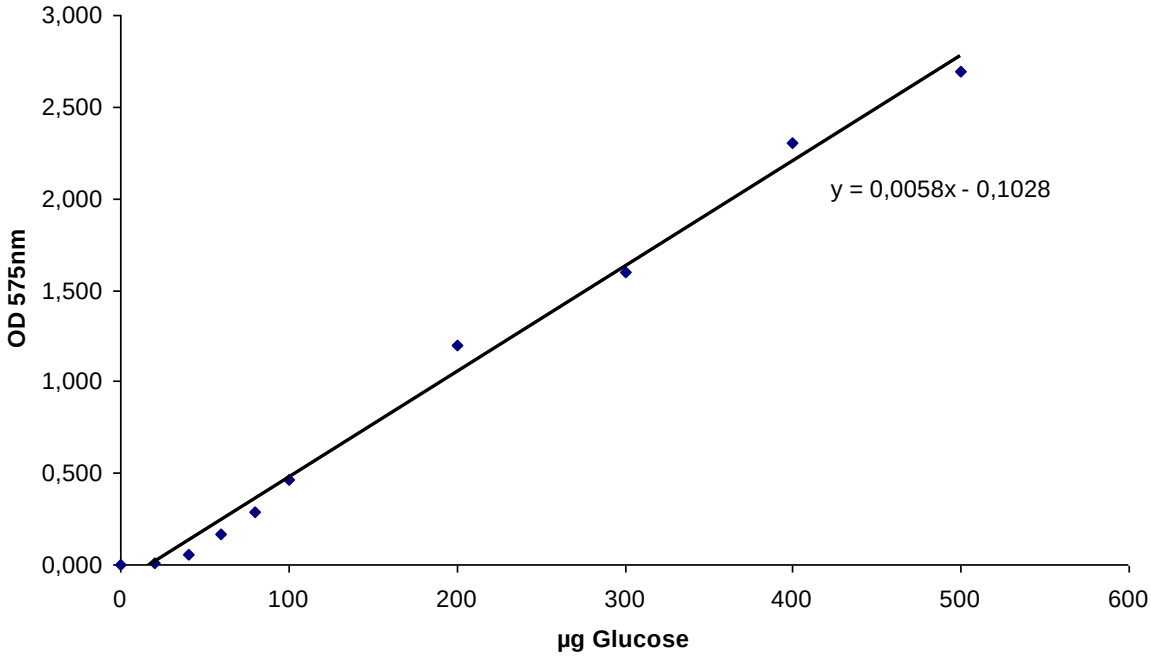
```
1684  ATC GGA ACA TGT GTC AAC TAT CCG TTT TAC AAC AAT TCA GAT CCA ACC TAC
        I   G   T   C   V   N   Y   P   F   Y   N   N   S   D   P   T   Y

1735  AAC AGC ATT TTG CAA AGA GAA TTT TCA ATG GTT GTA TGT GAA AAT GAA ATG
        N   S   I   L   Q   R   E   F   S   M   V   V   C   E   N   E   M

1786  AAG TTT GAT GCT TTG CAG CCG AGA CAA AAC GTT TTT GAT TTT TCG AAA GGA
        K   F   D   A   L   Q   P   R   Q   N   V   F   D   F   S   K   G

1837  GAC CAG TTG CTT GCT TTT GCA GAA AGA AAC GGT ATG CAG ATG AGG GGA CAT
        D   Q   L   L   A   F   A   E   R   N   G   M   Q   M   R   G   H

1888  ACG TTG ATT TGG CAC AAT CAA AAC CCG TCA TGG CTT ACA AAC GGT AAC TGG
        T   L   I   W   H   N   Q   N   P   S   W   L   T   N   G   N   W

1939  AAC CGG GAT TCG CTG CTT GCG GTA ATG AAA AAT CAC ATT ACC ACT GTT ATG
        N   R   D   S   L   L   A   V   M   K   N   H   I   T   T   V   M

1990  ACC CAT TAC AAA GGT AAA ATT GTT GAG TGG GAT GTG GCA AAC GAA TGT ATG
        T   H   Y   K   G   K   I   V   E   W   D   V   A   N   E   C   M

2041  GAT GAT TCC GGC AAC GGC TTA AGA AGC AGC ATA TGG AGA AAT GTA ATC GGT
        D   D   S   G   N   G   L   R   S   S   I   W   R   N   V   I   G

2092  CAG GAC TAC CTT GAC TAT GCT TTC AGG TAT GCA AGA GAA GCA GAT CCC GAT
        Q   D   Y   L   D   Y   A   F   R   Y   A   R   E   A   D   P   D

2143  GCA CTT CTT TTC TAC AAT GAT TAT AAT ATT GAA GAC TTG GGT CCA AAG TCC
        A   L   L   F   Y   N   D   Y   N   I   E   D   L   G   P   K   S

2194  AAT GCG GTA TTT AAC ATG ATT AAA AGT ATG AAG GAA AGA GGT GTG CCG ATT
        N   A   V   F   N   M   I   K   S   M   K   E   R   G   V   P   I

2245  GAC GGA GTA GGA TTC CAA TGC CAC TTT ATC AAT GGA ATG AGC CCC GAG TAC
        D   G   V   G   F   Q   C   H   F   I   N   G   M   S   P   E   Y

2296  CTT GCC AGC ATT GAT CAA AAT ATT AAG AGA TAT GCG GAA ATA GGC GTT ATA
        L   A   S   I   D   Q   N   I   K   R   Y   A   E   I   G   V   I

2347  GTA TCC TTT ACC GAA ATA GAT ATA CGC ATA CCT CAG TCG GAA AAC CCG GCA
        V   S   F   T   E   I   D   I   R   I   P   Q   S   E   N   P   A

2398  ACT GCA TTC CAG GTA CAG GCA AAC AAC TAT AAG GAA CTT ATG AAA ATT TGT
        T   A   F   Q   V   Q   A   N   N   Y   K   E   L   M   K   I   C

2449  CTG GCA AAC CCC AAT TGC AAT ACC TTT GTA ATG TGG GGA TTC ACA GAT AAA
        L   A   N   P   N   C   N   T   F   V   M   W   G   F   T   D   K

2500  TAC ACA TGG ATT CCG GGA ACT TTC CCA GGA TAT GGC AAT CCA TTG ATT TAT
        Y   T   W   I   P   G   T   F   P   G   Y   G   N   P   L   I   Y

2551  GAC AGC AAT TAC AAT CCG AAA CCG GCA TAC AAT GCA ATA AAG GAA GCT CTT
        D   S   N   Y   N   P   K   P   A   Y   N   A   I   K   E   A   L
                    •
2602  ATG GGC TAT TGA TAA TTC CGA AAA GCT GAG CAG ATA ATG ATG CCG TAA AGC
        M   G   Y   -   -   F   R   K   A   E   Q   I   M   M   P   -   S

2653  CGG CTT CTG AAT TAA GAG CCG GCT TTA CGG AGA TAT ACT TTT TAC GGC AGA
        R   L   L   N   -   E   P   A   L   R   R   Y   T   F   Y   G   R
```

Regression curves

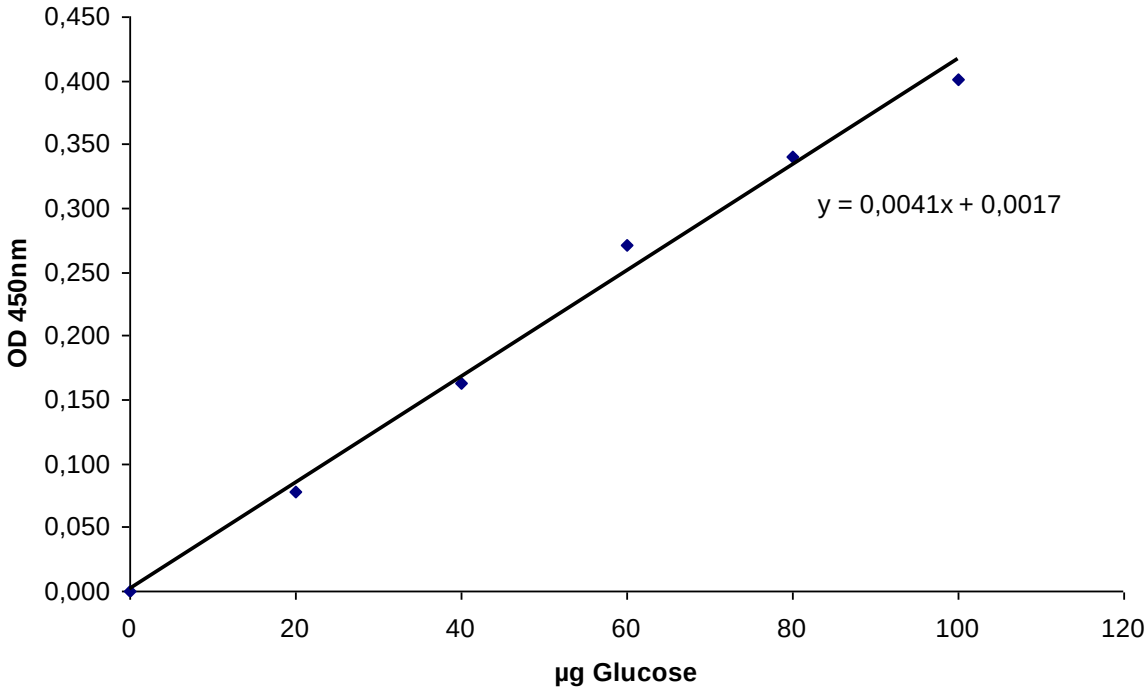
Bradford-Assay:



DNSA-test:

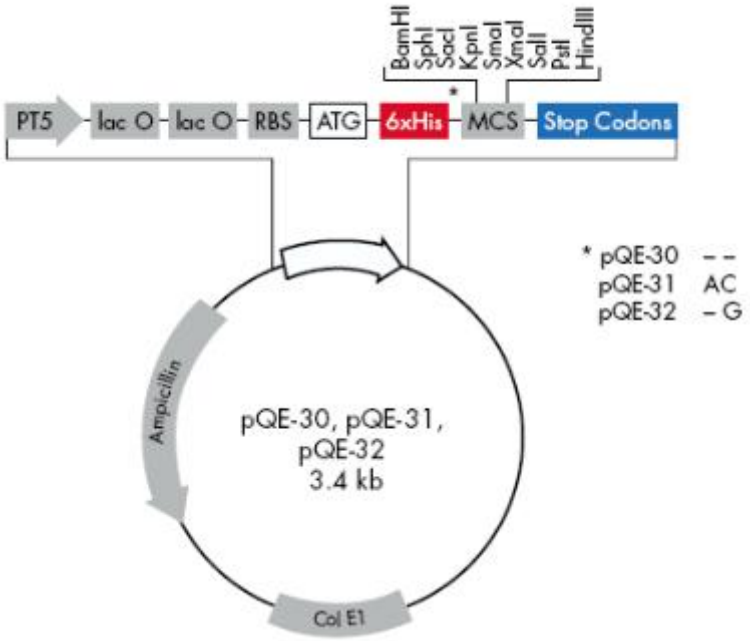


PGO-test:

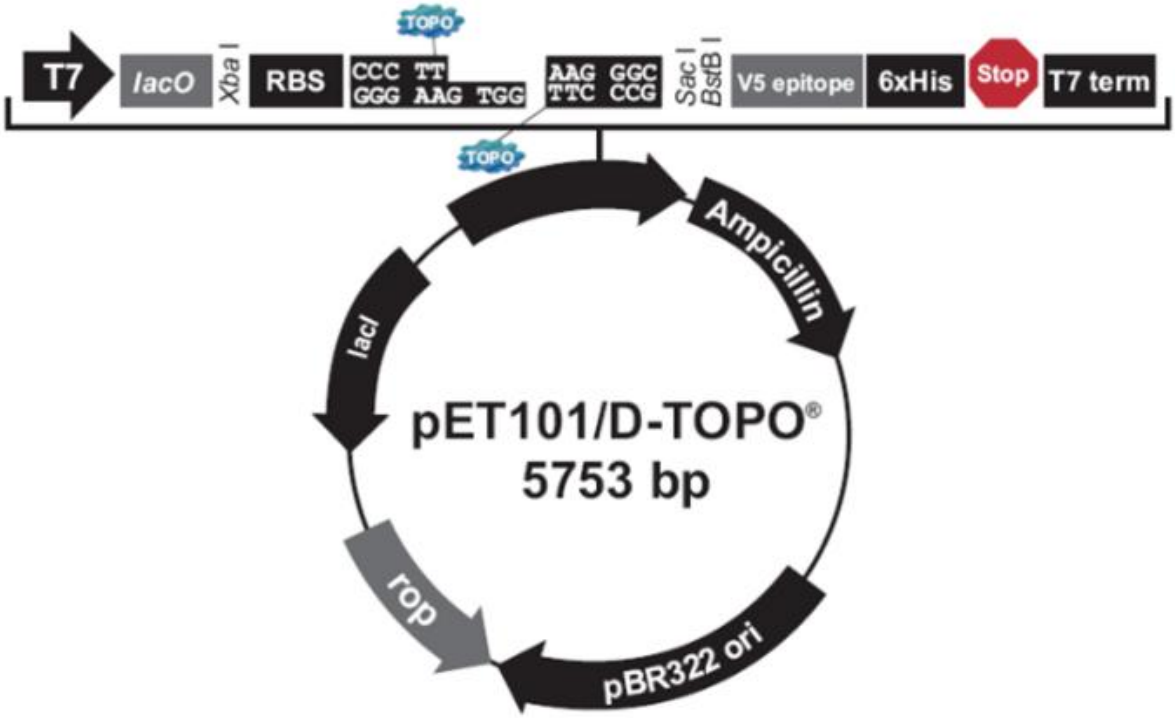


Vectors and plasmids

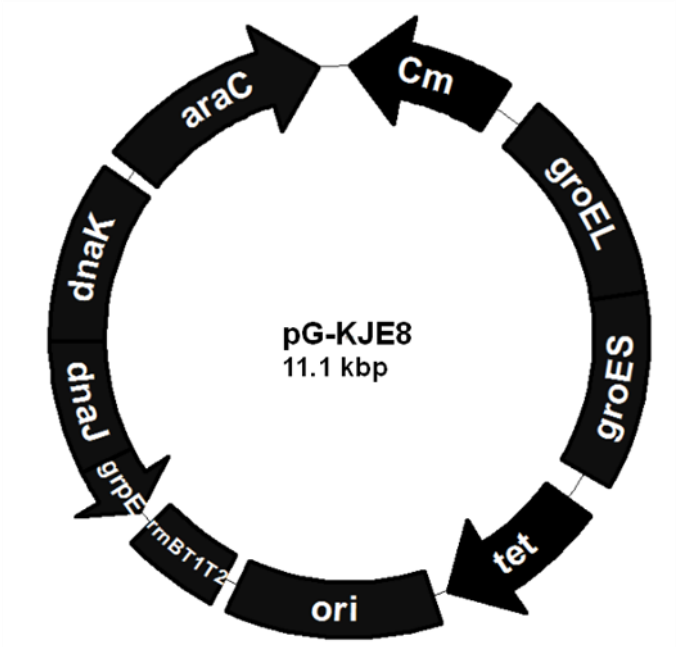
pQE 30-32:



Champion pET101/D-TOPO:



pG-KJE8:



Danksagung

Diese Arbeit wurde am Lehrstuhl für Mikrobiologie der Technischen Universität München von 2007 bis 2010 unter der Leitung von Prof. Dr, W. Liebl angefertigt und im Rahmen des Piltoprojekt Lignocellulose-Bioraffinerie – Projekt Enzymtechnologie – als Verbundvorhaben der Partner Universität Göttingen, Universität Hamburg, Technische Universität München, Merck KGaA finanziell unterstützt.

Herrn Prof. Dr. Wolfgang Liebl gilt mein herzlicher Dank für sein fortwährendes Interesse an dieser Arbeit und für die umfassende Unterstützung, die es möglich machte, diese Arbeit an seinem Lehrstuhl durchzuführen.

Herrn Dr. Wolfgang Schwarz danke ich für die interessierte Begleitung und Betreuung des Themas und die Hilfe in allen erdenklichen Fragestellungen.

Herrn Dr. Vladimir Zverlov gilt mein besonderer Dank für die unzähligen Tipps und Tricks im Umgang mit molekularbiologischen Techniken.

Herrn Philipp Schulte danke ich für sein großes Engangement bzgl. des Existsantrags, für sein Interesse an diesem Thema und für die abertausende, löchernde Fragen.

Herrn Prof. Dr. Christian Becker vom Lehrstuhl für Chemie danke ich für die Synthese und Bereitstellung des Polymers, welches die Basis für unsere weiteren Forschungsarbeiten darstellt.

Frau Johanna Häusler und Frau Angela Glaeser danke ich herzlich für die Klonierung, Expression und Aufreinigung einiger hydrolytischen Komponenten.

Herrn Dr. Angel Angelov danke ich für die Überlassung des Plasmids pMK-18

Danke an Anja Junker für die unzähligen Watten-Partien, und dafür dass Du so bist wie Du bist…

Danke an Herrn Dr. Martin Schlapschy vom Lehrstuhl für Biologische Chemie für die Tipps bzgl. der Anbindungschemie von Proteinen an Oberflächen

Danke an Herrn Dr. Matthias Schnabelrauch von der Innovent e. V. Jena für die Bereitstellung der Nanopartikel für die ersten Schuß-ins-Blaue-Versuche.

Danke an alle Angehörige des Legrstuhls für Mikrobiologie der TU München für das angenehme Arbeitsklima und insbesondere gilt mein Dank den Technischen Angestellten, ohne die der tägliche Betrieb an unserem Lehrstuhl nicht möglich wäre.

Danke an meine Familie und Freunde, die mich während meiner Zeit hier am Lehrstuhl immer unterstützt haben.

Curriculum vitae

PERSONAL DEATAILS:

- Date of birth: 14.10.1979
- Place of birth: Wertheim, Germany
- Marital status: Unmarried
- Nationality: German
- Parents: Dipl. Physicist Wolfgang Krauß,
  Elke Kessler-Krauß

SCHOOLAR TRAINING:

1986 – 1990	Primary school Muennerstadt, Bavaria
1990 – 1999	Johann-Philipp-von-Schoenborn Gymnasium Muennerstadt
1999	General entrance requirement for higher education (examination subjects: Biology, Physics, English, Economy), grade: 2.6
1999 – 2000	Military service at 8. GebStFmLehr-Batallion

UNIVERSITY TRAINING:

2000 – 2005	Study of Biology (Diploma) at the Bayerischen Julius-Maximilians-Universität Würzburg
June 2005	Diploma Genetic: 1.0 Microbiology: 2.0 Biotechnology: 1.7
July 2005 – May 2006	Diploma thesis at the Max-Planck-Institute of Biophysics Frankfurt/Main Topic: Electrophysiological measurements of the interactions of lipophilic aniones with biological and artificial lipidmembranes, grade: 1.0

WORK PLACEMENTS WHILE STUDYING:

F1 – work placement microbiology: Basis for applied microbiology, physiology of microorganism, analyse of differential gene expression, EST genome sequencing, insertion mutagenesis, effort of different reporter systems

F1 – work placement genetic: Enhancer-trap method, molecular genetic, GAL4-UAS-system, immunehistochemistry and Western-Blot, confocale-laser-scanning microscopy, fluorescence spectroscopy, PCR

F1 – work placement biotechnology: Basis of waste water treatment, electrical fieldeffects on biological structures (cell fusion, electropermeabilisation and cell manipulation), membrane transport processes in artificial lipidmembranes, immobilization of mice lymphocytes with alginate, purification of proteins

F2 – and special work placement biotechnology: Patch-Clamp-technique on *Valonia utricularis*, TEVC-method on *Xenopus laevis* oocytes

PROFESSIONAL EXPERIENCE:

Jan. 2007 – Jul. 2010 PhD thesis at the Department of Microbiology, Technical University Munich (TUM), Germany
Advisor Prof. Dr. Wolfgang Liebl
Title: Artificial cellulosome: *in vitro* reconstitution of the cellulosome of *Clostridium thermocellum*, grade: magna cum laude

TEACHING ACTICTIVITY:

Supervision of several bachelor and master students
Supervision of practical trainings for undergraduate and graduate students
Supervisor of seminar for microbiology for undergraduate students

KNOWLEDGES AND SKILLS:

Computer skills: Knowledges in MS-Office-Applications (Word, Excel, Powerpoint), Origin, Corel-Draw, MS-Operating system (Windows 3.11 to XP), Clone-Manager, Bio-Edit
Foreign languages: English and French fluently, great Latinum
Other skills: Fast perception, creativity, team spirit, experimental expertise, independent working

HOBBYS AND INTERESTS:

Sport: Hornussen (team), Swimming, Soccer, Basketball, Jogging, Snowboarding
Spare time: Music, reading, nature, culture