Aleksandra Mitrović

Discovery and Optimization of Fungal Lignocellulolytic Enzymes

Doctoral Thesis

Supervisor: Ao.Univ.-Prof. Mag. Dr.rer.nat. Anton Glieder

Institute of Molecular Biotechnology / Austrian Center for Industrial Biocatalysis

Graz University of Technology, Austria

2013

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am 19.09.2013 ……
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, 19.09.2013 ……
(Signature)

Acknowledgements

My honest gratitude goes to:

Prof. Toni Glieder, for offering me the great opportunity to enter the world of applied science in such a beautiful and inspiring environment;

Sandra Abad, for being my first science-guru, for all the things and tricks she taught me, for making lab work so colorful, for the help and understanding; Harald Pichler, for never-ever telling me "I don't have time", whenever I had a question or a dilemma, as well as for his endless calmness and patience; Christoph Reisinger and Georg Schirrmacher for the pleasant atmosphere within the project and smooth cooperation throughout the years;

the 3rd floor crew, for being so patient and helpful when I was just a "young grasshopper", and to Thorsten Bachler, for showing me his "secret" competent E. coli stash and generously sharing all his stock solutions with me;

the 2nd floor crew, for the cozy atmosphere and fun late-evening chats, and especially to Andrea Camattari, for all the constructive discussions and encouragements when I needed them most, and Andrea Mellitzer, for being always available and an awesome teammate.

In addition I'd like to thank Mihi and Matko for the graphical support throughout the thesis; Julia and Nikola for the grammar and spell checks; Ivana for never getting tired of my stories, and Viktorija for the help and understanding.

I owe a very special thanks to Raimund, for inspiring me, encouraging me, for having my back, for all the oh-so needed chill out moments, for the love and caring.

Above all, I am grateful to my family. I thank my three wonderful siblings, for all that they mean to me, for always cheering me up, and for being so close despite the distance. My deepest gratitude goes to my parents: for raising me to be who I am, for showing me learning is fun and teaching me to cherish knowledge, for being inexhaustibly devoted, loving and supportive.

Abstract / Kurzzusammenfassung

There is a strong need for development of so-called 2nd generation biofuels, which are produced of non-edible plant biomass. This biomass is mostly composed of cellulose and hemicelluloses (e.g. xylan). These sugar polymers are highly recalcitrant to mild degradation without other chemical modification; however there are several approaches for their degradation to fermentable sugars, with biocatalysis being a promising choice. Enzymatic degradation of (hemi)cellulose requires a mix of enzymes specific for various chemical bonds within these complex biopolymers. For making the enzymatic degradation commercially applicable new enzymes are constantly being discovered from nature, and also known enzymes are often engineered to better suit the industrial processes.

Within this thesis, first a review of available techniques for enzyme discovery and engineering is given. Further, the discovery of three novel fungal xylanases following a novel concept of functional cDNA library screening is presented. This was achieved by activity screening of a fungal transcriptome library expressed in the fission yeast Schizosaccharomyces pombe. The second part of this thesis focuses on engineering of endoglucanase Cel7B from Hypocrea pseudokoningii towards higher thermostability. The used approach included random mutagenesis, as well as structure-guided design. Final improvement of 10°C in terms of melting temperatures was achieved.

Es besteht ein großer Bedarf für die Entwicklung von so genannten Biokraftstoffen der zweiten Generation, welche aus nicht-essbaren Pflanzen hergestellt werden. Jene Biomasse setzt sich hauptsächlich aus Cellulose und Hemicellulose (z.B. Xylan) zusammen. Diese Zucker Polymere sind sehr schwer abbaubar, ohne dabei chemische Veränderungen der Produkte zu verursachen;, dennoch gibt es einige Ansätze zur Spaltung in fermentierbare Zucker. Biokatalyse bietet eine der vielversprechendsten Möglichkeiten. Der enzymatische Abbau von Cellulose bzw. Hemicellulose erfordert eine Mischung an Enzymen, welche spezifisch auf die verschiedenen chemischen Bindungen dieser komplexen Biopolymere wirken. Um enzymatischen Abbau kommerziell zu nutzen, werden zum einen ständig neue Enzyme aus der Natur entdeckt, und zum anderen bereits vorhandene und bekannte Enzyme für den industriellen Prozess optimiert.

Im Rahmen dieser Arbeit wird zunächst ein Überblick der verfügbaren Techniken zur Erforschung und Optimierung von Enzymen gegeben. Ebenso wird die Entdeckung von drei neuen Pilz-Xylanasen mit Hilfe eines neuartigen funktionellen Screening Ansatzes präsentiert. Dies wurde durch funktionelles Screening einer pilzlichen Transkriptom-Bibliothek erreicht, wobei die Spalthefe Schizosaccharomyces pombe als Wirt verwendet wurde. Der zweite Teil dieser Arbeit konzentriert sich auf die Optimierung von Endoglucanase Cel7B von Hypocrea pseudokoningii, um eine höhere Thermostabilität zu erreichen. Für diesen Ansatz wurden sowohl Zufallsmutagenese als auch Struktur basierende Informationen verwendet. Schlussendlich wurde eine Verbesserung der Schmelztemperatur um 10°C erreicht.

Table of Content

1. Introduction

"If engineering is the application of science for human benefit than the engineer must be a student not only of the application of science, but of human benefit as well."

John M. Prausnitz

Modern world developed high dependency on a finite source of energy - the fossil fuels, comprising coal, petroleum, and natural gas. For example, the transportation sector (including airlines, railroads and trucking companies) is more than 96% dependent on petroleum [1]. Not only is this source of energy finite, but also has other serious drawbacks. Its high emission of green house gases (GHG) is altering the global climate [2], [3]. The mentioned transportation sector consumes approximately two-thirds of extracted crude oil, and accounts for approximately one-third of greenhouse gas emissions in the world [1]. In addition, according to the International Tanker Owners Pollution Federation (ITOPF), since the late 1960s there have been close to 10000 accidental oil spills, ranging from minor ones to the ones causing huge environmental disasters. Due to the technology and logistic developments in the oil shipping industry, only 5% of the large oil spills happened in the 2000s, indicating a dramatic decline. Nevertheless, oil extraction is moving into ever-deeper water and tempestuous seas, and environmental, as well as socio-economic, risks are eminent [4], [5]. Altogether, rapid depletion of fossil fuel reserves and accumulating environmental imprints of its exploitation gave rise to a search for a sustainable alternative to global petroleum-addiction [6].

The most common forms of renewable energy include hydro-, wind-, and solar-power, and biomass. In this context, biomass deserves an exceptional focus, as the only known, large-scale, renewable resource that can be converted into the liquid fuels that are so well established in transportation [7], [8]. Vegetable oils can be used as raw material for biodiesel production, but they are not nearly as abundant as general biomass [9]. Liquid and gas fuels produced from biomass are referred to as the biofuels, and they include bioethanol, biomethanol, biobutanol, biodiesel, and biohydrogen. Among the biofuels, bioethanol has received considerable attention in the transportation sector because of simple manufacturing and its utility as an octane booster, fuel additive, and even as pure fuel [10]. In fact, the first mass-produced car in history, the legendary Ford Model T from 1908, was designed to run on bioethanol, or as Henry Ford labeled it "the fuel of the future" [11], [12]. Since then, interest in

bioethanol was fluctuating, but with current environmental, political, and economical circumstances, it is again in the spotlight.

The first generation of biofuels to emerge on the market was derived mainly from food crops. The long-term success of this "trial run" to achieve independency of fossil fuels is limited by several factors. First, it raised many debates due to competitiveness with land and water utilized for food production, leading to increased food prices [13]. Second, although one of the objectives of switching to biofuels was reducing GHG emissions and being eco-friendly, there are serious indications that the first generation of fuels failed at achieving it. Beside the obvious devastating effect on wildlife, converting natural ecosystems to food crops for biofuel production also releases 17 to 420 times more CO_2 than the expected annual reduction by using biofuels produced from that land [14], [15]. To avoid this paradoxical effect and get back on the right track, a different approach was needed.

Second generation of biofuels is produced either from lignocellulosic waste biomass (crop and forest residues, wood process wastes, organic fraction of municipal solid wastes) or from low-maintenance perennial plants grown on degraded and abandoned agricultural lands. With practically no additional land requirements, this approach leaves the natural depot of organic carbon in soil and vegetation intact, and offers immediate and sustainable advantages in environmental terms [16]. In addition, taking into account energy input for agricultural processes and biorefining, second generation biofuels offer higher energy output then the food-crop based biofuels [17]. To sum up, bioethanol produced from lignocellulosic biomass is a promising alternative since the raw materials are highly abundant, do not directly compete with food production, they are environment-friendly, and demand less funding and energy investments [18], [8]. Also, sustainable land use harvesting a part of the biomass for fuels and chemical production while leaving back a significant fraction of the biomass in the soil is feasible.

Although the precise composition of lignocellulose strongly depends on the type and origin of the particular plant biomass, lignocellulose in general consists of three types of polymers: cellulose (35-50%), hemicellulose (20-35%), and lignin (5-30%) [19]. Native cellulose is a linear polymer of about 10000 D-glucopyranosyl units per chain, linked together by β-1,4-glycosidic bonds. Due to the 180°-rotation between adjacent glycosyl units, each side of the cellulose chain has the same number of hydroxyl groups, thereby leading to a symmetrical and linear structure. Around 30 individual cellulose chains stabilized by strong intermolecular hydrogen bonds between hydroxyl groups are assembled into larger units called elementary fibrils, which are further arranged into microfibrils. Microfibrils are further assembled into macrofibrils, also called "fibers". The parallel alignment and the extensive number of hydrogen bonds cause a tightly packed crystalline structure, which completely prevents penetration of

any molecules, including water. However, microfibrils also contain amorphous regions that are potential points for chemical and enzymatic attacks [20]. In order to stabilize its supramolecular structure, cellulose fibrils are embedded in a matrix of hemicellulose and lignin tightened by strong physico-chemical bonds [21].

Hemicellulose is a short, highly branched heteropolymer of pentoses (D-xylose and L-arabinose), hexoses (D-galactose, D-glucose, and D-mannose) and glucuronic acid, whose exact composition is highly dependent on the plant source. Hemicelluloses are often referred to as xylans, mannans or galactans depending upon the predominant sugars. Its monomers are linked through 1.3; 1,6, and 1,4 glycosidic bonds and are often acetylated. Hemicellulose forms a loose hydrophilic structure acting as glue between cellulose and lignin. Lignin is a highly branched polyaromatic compound consisting of phenylpropanoid units, mainly paracoumaryl, coniferyl, and sinapyl alcohols. It forms covalent bonds with hemicellulose, and connects the individual complexes of microfibrils and hemicellulose. Lignin is hydrophobic and highly resistant to chemical and biological degradation. Mutual association between cellulose, hemicellulose, and lignin, combined with tight packing of cellulose fibers, lead to high mechanical stability of lignocellulose, and to high resistance against chemical and enzymatic degradation [6], [21].

This natural recalcitrance of lignocellulose is the cause of high costs of its conversion [6], and the reason why 2nd generation biofuels still face major constraints to their commercial deployment. Despite all the potential of lignocellulose, pilot-plants producing 2nd generation biofuels supply less than 0.1% of world biofuel market. This is partially due to the high investment risks associated with complex evaluations of their economics [22], [23], [24], [25], although in the long run biomass markets were expected to bring much more economic/political stability as opposed to the oil price fluctuations [26]. So far the lignocellulose processing technologies are relatively immature, and consume 70% of the overall production cost (while in crude oil refining the same proportion of costs is designated for raw materials). This gives much room for reducing the costs by increasing process efficiency [9], [16].

Ethanol production from lignocellulosic biomass comprises the following main steps: pretreatment of lignocellulose, hydrolysis of cellulose and hemicellulose, sugar fermentation, separation of lignin residue, and product recovery [21]. Hydrolyzing lignocellulose to fermentable monosaccharides is problematic due to its tight packing and strong association of its main components. Owing to these structural characteristics, pretreatment is an essential step with the aim to break down the lignin structure, disrupt the crystalline structure of cellulose, and ultimately enhance enzymes accessibility to the cellulose. Among the different methods of pretreatment, chemical and thermochemical are the most effective ones for industrial applications [8]. Besides being a crucial step in the biological conversion to ethanol, biomass

pretreatment has been described as the most expensive step in the conversion of lignocellulose to ethanol excluding the raw material price [1], [27], [28].

Despite the fact that the pretreatment is the most costly step, there are indications that optimization of the subsequent hydrolysis step might play the crucial role of reducing overall process costs [29], [30]. Lignocellulose degradation can be done either thermo-chemically or by biocatalytic approach, in which enzymes or micro-organisms are used to hydrolyze (hemi)cellulose to sugars. The biocatalytic route is less mature, and thus offers a greater potential for cost reduction [16]. Also, in comparison to the thermo-chemical degradation, biocatalytic conversion has high selectivity, exhibits higher conversion efficiencies, and it uses mild conditions and thus has low equipment costs [6], [13]. The mentioned selectivity of the biocatalytic conversion leads to another key difference from the thermo-chemical route. Enzymatic hydrolysis yields lignin as a residue, which is further used to supply heat and power to the process itself, resulting in higher net-energy output [31]. Although intensive research since the 1980s brought significant process improvements, the enzymatic hydrolysis of lignocellulose step remains as a major techno-economic bottleneck in biomass-to-ethanol conversion [32], [21]. Much remains to be done in terms of improving the efficacy of enzymes and achieving high biofuel yields at low enzyme loadings, lowering enzyme production costs, and improving overall process integration [33], [13].

Particularly fruitful target for reaching economic viability of 2nd generation biofuels could be the cellulose-degrading enzymes, including endo-1,4-β-D-glucanases, cellobiohydrolases, and β-glucanases (**Figure 1**). In 2006, costs of cellulases needed for production of one unit of ethanol by lignocellulose saccharification were around 4 to 10 times higher than costs for amylases used in the starch-to-ethanol process [9], while ethanol production from sugarcane and molasses entirely bypasses this cost [10]. This emphasizes the need for development of ever-better cellulases, in terms of high specific activities, high turn over rates, increased thermostability, decreased inhibition by reaction products, synergism among the different enzymes, etc. [34], [35]. While costs for cellulase production might be decreased by over-expression techniques [36], increasing their performance is often fulfilled by protein engineering approaches [37], [6], [13].

Increasing enzyme thermostability is a promising step towards commercializing 2nd generation biofuels [35], especially when considering the growing trend towards integrated process configurations which significantly lower overall costs [19], [38], [39], [21]. Since most current biomass pretreatment procedures take place at high temperature, thermostable enzymes are of paramount importance to avoid unnecessary energy input into the process by intensive cooling before enzymatic hydrolysis and for increasing flexibility of such biorefinery process configurations. Other important advantages of thermostable enzymes include allowing extended

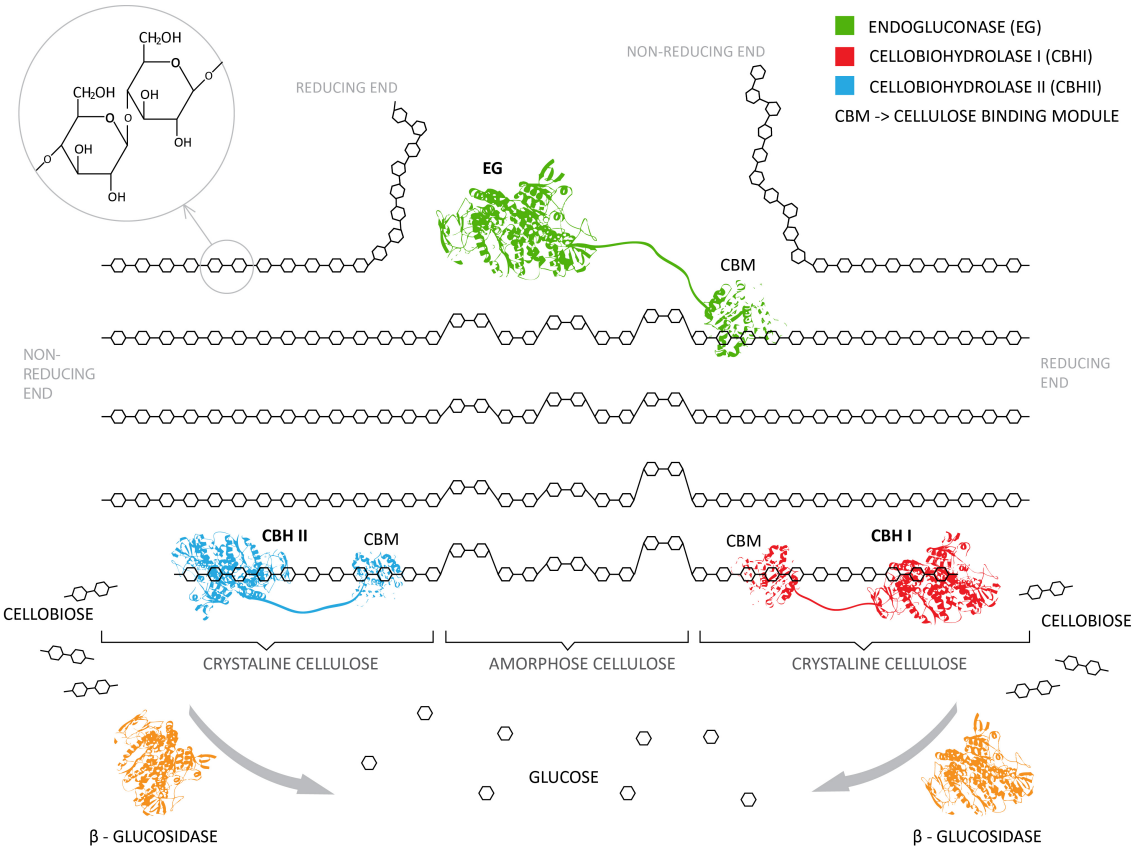

Figure 1 – Schematic overview of enzymatic degradation of cellulose.

hydrolysis times and decreased enzyme amounts, being more compatible with non-enzymatic pretreatment processes and eliminating/decreasing associated cooling costs, etc. All these benefits surely make the overall process more economically attractive [40].

Further process development towards the economic competitiveness of lignocellulosic ethanol, uncompromisingly demands effective utilization of all three biomass components. Structural heterogeneity is a hallmark of hemicelluloses, and its effective hydrolysis requires a spectrum of enzymes to be employed [16]. One of the most important hemicellulases is endo-1,4-β-D-xylanase (endoxylanase), which cleaves glycosidic bonds in the xylan backbone. The positive effect of xylanases on overall yield of sugars by lignocellulose hydrolysis has experimentally been demonstrated by several authors [41], [30], [42]. The mechanisms by which hemicellulases facilitate cellulose hydrolysis include exposing the cellulose fibers and making them more accessible, as well as removal of xylo-oligosaccharides, which inhibit cellulase enzyme activity [43]. In addition, the importance of matching a right type of xylanase with the particular biomass used has been reported [41]. This arises from the fact that xylan structure varies greatly depending on the source. In this context, current efforts have

been addressed to mining microbial diversity to obtain novel enzymes for effective hemicellulose degradation [44], [8], [6]. Filamentous fungi are an especially promising source of novel xylanases, as they excrete the enzymes into the medium and their enzyme levels are much higher than those of yeasts and bacteria [45]. Discovering new enzymes is especially important, as protein engineering simply does not always offer enough sequence space for required improvements [7]. On the other hand, many extraordinary enzymes are protected from public disclosure by patents, which further builds the need for novel enzymatic sequences [1]. In addition, recent discoveries such as cellulose cleaving monooxygenases demonstrated that not all enzymes involved in efficient lignocellulose degradation are known yet [46].

There's room for improvement in almost every facet of biofuel production from lignocellulose [9]. New and advanced enzymes continue to be developed in these purposes and this thesis contributes to this field. Within this thesis, a set of novel xylanases was discovered by activity screening of a transcriptome originating from a wood rotting fungus *Fomes fomentarius*. The applied method presents a unique but promising approach for enzyme discovery, by employing *Schizosaccharomyces pombe* as an expression host for a eukaryotic cDNA library. In addition, endoglucanase Cel7B from *Hypocrea pseudokoningii* was improved in terms of thermostability, by combining random mutagenesis and structure guided design.

2. New Enzymes: Complementing Natural Diversity with Laboratory-Based Methods*

Efficient and sustainable biocatalytic processes rely on cheap enzymes and their capability to perform selective chemical reactions. A major limitation of industrial process development is the lack of robust enzymes with novel functions or adapted properties for industrial application on non-natural substrates in a non-natural environment. Therefore, a strong focus in industrial biotechnology and biocatalysis is on exploiting natural and laboratory-generated diversity of enzymes to reduce non-selective side activities and to drive the spectrum of biocatalytic processes into new directions. In this article we outline basic methods for the exploitation of nature's diversity in enzymatic functions and focus on new concepts of laboratory-based methods that extend the range of available enzymatic functions.

2.1. Enzymes – Sustainable Solutions for Biotechnology

Major efforts towards sustainable industrial production processes and reduced CO_2 footprints in recent years as well as economical advantages have led to an increased application of enzymes in industry. Enzymes allow the development of environmentally friendly, highly efficient, energy- and resource-saving processes that are selective and specific under mild conditions and often preferable over chemical catalysis. However, industrial processes and the compatibility of biocatalytic steps with other unit operations still often require conditions that are harsh for native enzymes. Therefore, improvement of stability and activity of existing enzymes and the identification of more appropriate novel enzymes are still indispensable for a successful replacement of conventional chemical processes by biocatalytic processes. Consequently, a major goal is to meet the requirements of the industry by providing novel and adapted enzymatic functions from natural diversity or laboratory-based methods (**Figure 2**).

2.2. Natural Diversity

2.2.1. Microbial Diversity

With their rapid proliferation times and 4 billion years of evolution under often extreme selective pressure microorganisms formed a fantastic natural resource of enzymatic diversity. In addition, the emerging production of atmospheric oxygen by photosynthetic bacteria some 2.5 billion years ago has been a key step for the evolution of oxygen-dependent redox-enzymes, such as cytochromes, oxidases, dehydrogenases, and hydroxylases. Most of these have found broad application in today's enzyme-based biotechnology, and especially in biocatalysis.

* Mitrovic, A. Glieder, K. Flicker; Chemistry Today, 2011, Vol.:29, No.:2

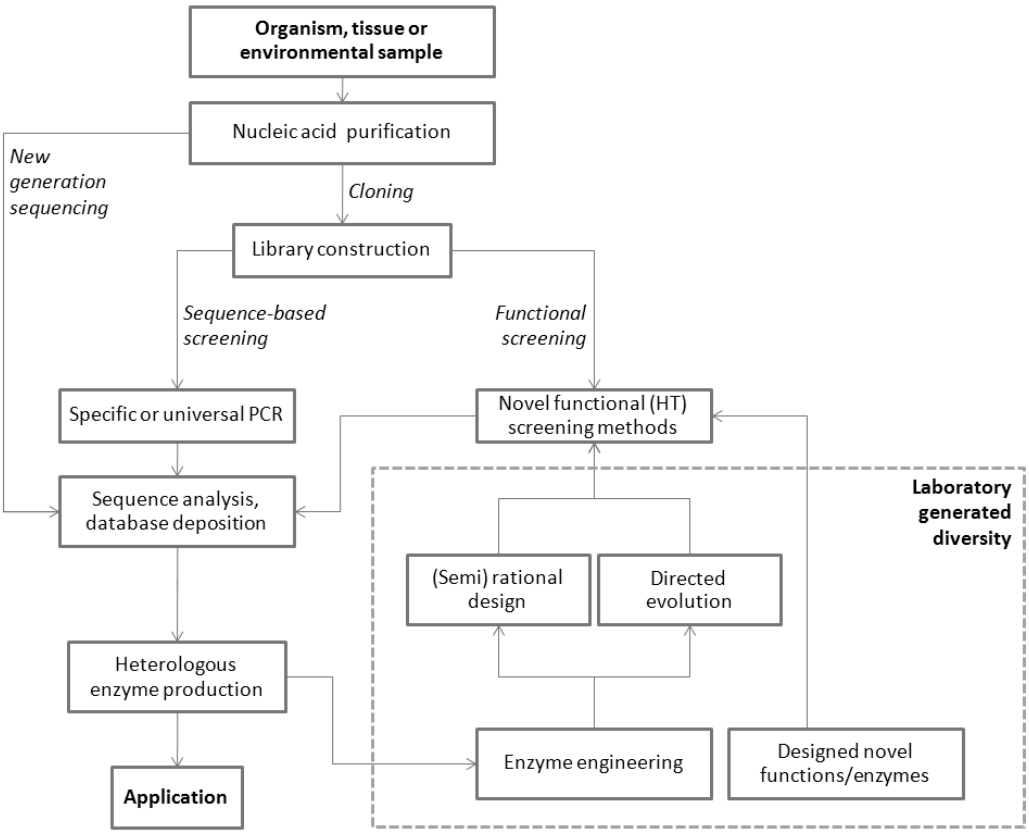

Figure 2 - Novel enzymes from natural diversity and laboratory-based approaches.

Direct microbial screening has been the classical approach for enzyme discovery for many years. However, the vast majority of (novel) enzymatic functions is not accessible this way; in fact less than 1% of the microorganisms are cultivable in the lab [47]. This is especially true for microorganisms from "extreme" habitats, endosymbionts and other niche-specialized communities from which interesting novel enzymes can be expected. Due to recent technological breakthroughs in DNA and RNA sequencing [48] metagenomics has become a major method for enzyme discovery. Metagenomics is cultivation-independent and combines functional- and/or sequence-based (*in silico*) approaches to directly access the genetic diversity of complex microbial communities [49]. However, efficient heterologous expression of new genes from metagenomic libraries is often challenging [50].

2.2.2. Functional/Activity-Based Screening

The generation of metagenomic DNA libraries is an established method for function-based metagenomic gene discovery. *E. coli* is the most commonly used host for functional metagenomics, mostly because of its versatility as heterologous host and its good characterization on genetic and physiological level. However *E. coli* is often

insufficiently versatile and biased for protein expression and functional screening. In addition to transcriptional problems functional expression in *E. coli* may also be limited on translational and posttranslational level either by non-functional ribosome binding sites, adverse RNA secondary structures, rare codons in the target gene, missing cofactors and folding assistance by chaperons, or by the missing export and processing machinery required for proteins with signal sequences. Despite the proven usefulness of *E. coli* in metagenomics different host expression systems are required to overcome these limitations of current functional metagenomic screening. Alternative hosts are e.g. *Bacilli, Pseudomonades* and *Streptomycetes*. Eukaryotic expression systems have so far not been established or at least not published. Although mentioned in patent applications the proof-of-principle for efficient functional screening of eukaryotic genomic libraries in yeast still has to be demonstrated. The overall success of metagenomics to achieve real novelty in enzyme discovery has been evaluated in detail by Tuffin *et al.* [51]. In their evaluation they address function space (enzymatic function), sequence space (protein sequence), and topology space (protein structure). Based on their evaluation only a minimal number of novel structures has been identified through metagenomics. They also conclude that current functional screening has delivered novel sequences rather than novel activities which is a direct consequence of a limited availability of novel activity assays. Nevertheless these novel sequences with known functions are highly relevant for industrial applications in view of existing restrictions due to patent applications.

2.2.3. Metagenome Screening

Based on recent developments in next-generation sequencing complete metagenomes can easily be sequenced. Over 1.2 million new genes were discovered by Craig Venter's Sargasso Sea metagenome project [52]. Simple homology searches can certainly be useful in finding interesting enzyme coding genes. However, discovering enzymes with previously unknown activities is unlikely if metagenomes are only compared to already known genes. An option for finding novel enzymatic activities by this approach would be testing the new proteins for potential enzymatic promiscuity. In case of the mostly studied bacterial metagenomes site-based and content-based computational gene prediction is a possible alternative to the above mentioned comparative method. However, it still is not reliable for the prediction of eukaryotic genes because of their complex exon-intron structure. Probably just because of that a variety of computer programs for gene prediction is available. Apart from problems with reliable full length gene prediction also functional assignment for many novel genes is difficult as they show no homology to any known genes or only to other unnotated genes. In this context a combination of homology modelling and virtual screening was demonstrated to be very powerful for computational function discovery [53].

2.2.4. Sequence-Based Metagenomics

Sequence-based metagenomics can be used to directly access natural enzymatic diversity from extensive sequencing approaches but also without the need of a fully sequenced metagenome. Usually degenerate primers from consensus sequences of known enzymes are used for the amplification of target genes from metagenomic DNA. This approach, however, reduces the novelty in discovery as previously unknown enzymatic activities are unlikely to be identified. Regarding the novelty in sequences this approach is also limited although enzymes sharing less than 20% amino acid sequence identity can be retrieved [54]. Gene-specific PCR usually amplifies only a conserved region of a target gene and additional steps are needed to access the full length genes [55]. Furthermore, the combination of biases derived from primer and probe design and PCR and library bias often leads to an underrepresentation of low abundance sequences (of rare community members).

To overcome part of these problems subtractive hybridisation magnetic bead capture has been developed as a novel technique for the isolation of specific target genes from metagenomes [56]. In this approach multiple gene variants from a single amplification reaction can be retrieved, thereby accessing a larger fraction of the true microbial biodiversity. However, despite the expression bias, functional screening has proven superior in a recent comparison to sequence-based gene discovery methods [57].

2.2.5. Transcriptome Sequencing

Metatranscriptomics analyses the subset of genes transcribed in any given microbial community at a specific time and under specific conditions. Similar to metagenomics, sequencing and sequence-based or functional screening can be used for enzyme discovery. Although gene prediction is not required the problem of function assignment to novel sequences remains. Metatranscriptomics seems more powerful for the discovery of novel enzymes as full-length protein-encoding polyadenylated mRNA can be selectively isolated. Furthermore, eukaryotic protein-encoding genes, free of introns, are available this way. Due to the dynamic responses of transcriptomes towards environmental changes an “enrichment” of desired enzymatic activities can be achieved when microbial communities are intentionally exposed to defined conditions. This adaptability of transcriptomes is the basis for the generation of subtractive cDNA libraries which allows an even higher focus on interesting target enzymes. In this approach molecular biological tools are applied to make a transcriptome of an “induced” culture/community devoid of commonly expressed genes which are also present in a “non-induced” reference culture/community. An excellent PCR-based cDNA subtraction and normalization method termed suppression

subtractive hybridization was presented by Diatchenko *et al.* [58]. In this method cDNA subtraction was achieved by attaching inverted terminal repeats to the cDNA. These terminal repeats suppress the PCR amplification of undesirable sequences by head-loop formation and can therefore lead to a 1000-fold enrichment of rare and differentially expressed full-length sequences. In analogy to the metagenomic approach, metatranscriptomics has been successfully used to analyse entire microbial communities.

In addition to the selective isolation of protein-encoding genes and enrichment of functionally relevant genes from microorganisms transcriptome sequencing also provides access to a vast new diversity of genes and enzymes from higher eukaryotes. In contrast to expensive and time consuming sequencing of plant or animal genomes transcriptomics provides quick and relatively cheap access to many new coding sequences. For example, the ongoing “1000 *de novo* plant trancriptomes" project that was intitiated by the government of Alberta, Canada, aims at sequencing and assembling 1000 plant transcriptomes by 2012. This will hugely expand the knowledge of plant biodiversity, and deliver many new plant enzyme sequences.

2.3. Artificial Diversity

2.3.1. Directed Evolution

Many reviews give excellent summaries of well established directed evolution methods. Lutz and Patrick give a concise overview on current methods for directed evolution [59] and a detailed review is available by Yuan *et al.* [60]. Improving an existing enzymatic function or property by the accumulation of beneficial single mutations has been shown to be very effective. However, generation of new enzymatic functions, properties or the adaptation to new substrates may require multiple simultaneous mutations which are statistically unlikely to occur in combination in one individual protein variant. It is extremely difficult to practically achieve this with multiple rounds of directed evolution as the additivity and especially the synergistic action of single mutations cannot be foreseen. To aim for high mutation frequency also does not solve the problem as most of the generated variants would be inactive. As an alternative, the concept of substrate walking of Codexis & Merck offers new opportunities. It has recently been successfully applied to the API of the diabetes drug Januvia (Sitagliptin, **Figure 3A**) [61]. In this approach a transaminase which, initially, was not active on the target substrate was turned into an efficient enzyme for this target by sequential adaptation to similar substrate substructures (“substrate walking”) which linked the natural substrate to the desired target. In principle, this strategy of sequential adaptation to an artificial pathway by stepwise directed evolution was already applied by Chen and Zhao for the development of a novel

corticosterone binding activity in the α ligand-binding domain of the human estrogen receptor [62]. With their method termed "*in vitro* coevolution" Chen and Zhao have been able to evolve the α ligand-binding domain using testosterone and progesterone as hypothetical intermediates that provided a progressive structural bridge from 17β-estradiol to the final target, corticosterone.

As a general rule for extensive mutagenesis, higher initial stability provides higher mutational tolerance and gives more space for evolutionary adaptation in directed evolution. A somewhat different perspective on that phenomenon is that neutral mutations (in a more stable starting enzyme variant), although apparently being without effect, can generate more available adaptive pathways that lead to improved protein variants [63].

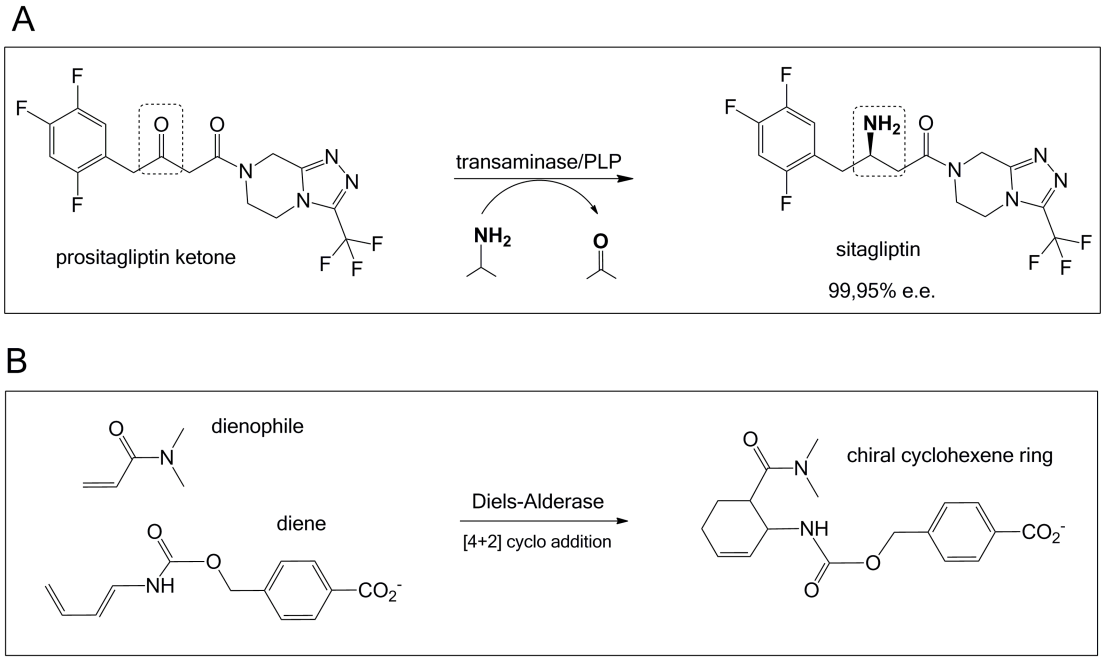

Figure 3 - Examples of novel enzymatic functions obtained by enzyme engineering: (A) shows the sitagliptin synthesis using a transaminase adapted to the nonnatural substrate by "substrate walking" (15), and (B) illustrates a stereselective bimolecular Diels-Alder reaction for which an enzyme was designed *de novo* by computational methods (27).

2.3.2. Recombination

An efficient way to access multiple mutations at a time while still obtaining folded and functional enzymes is homologous recombination – a concept to increase functional diversity that is frequently observed in nature. However, despite the functional relationship of genes they often show only low homology, which renders them unsuitable for homologous recombination. Therefore, also homology independent recombination methods were developed. As a recent example, the functional linkage

of unrelated proteins using the transposon-targeted domain insertion method (DIM) has been developed [64]. In their publication Edwards *et al.* used DIM to integrate a cytochrome b_{562} into the TEM-1 β-lactamase. By this integral protein fusion they showed that the structural change of the cytochrome caused by heme binding serves as a molecular switch for triggering β-lactamase activity.

It is known that recombination of (related) enzymes can lead to novel enzymes with improved properties at a high success rate, especially if the 3-dimensional structures of proteins are considered. An example for such a structure guided approach was developed in the group of F. H. Arnold. Applying the computer-based algorithm "SCHEMA" numerous highly functional chimaeric cellobiohydrolase II enzymes have been generated by structure-guided recombination [65].

2.3.3. Rational and Combinatorial Protein Engineering Techniques

Rational (structure- and mechanism-based) methods have the advantage that mutations are introduced based on available structural and functional information, which results in a concomitant increase of "useful" mutations. A broadly applicable (semi-) rational engineering concept considering structural information to increase enzyme stability is the B-Fit method by Manfred Reetz [66]. This method targets flexible structural elements (residues) of proteins that are identified by the high B-factors of their corresponding crystal structures. A similarily systematized and successful method of the Reetz lab to improve selectivities and activities became well known as ISM (iterative saturation mutagenesis) [67]. ISM combines rational design and combinatorial saturation mutagenesis with degenerated primers, thus providing on the one hand more focused libraries and on the other hand the full set of potential amino acid substitutions that would not be accessible with random approaches. Similar to these focused combinatorial libraries a combinatorial active-site saturation mutagenesis test (CASTing) was developed which specifically targets the active sites of enzymes and therefore can be used for the engineering of enzymatic properties such as enantio- or regioselectivity [68]. CASTing provides small and "smart" libraries as the amino acid substitutions can be varied according to the requirements by using degenerated codons. A combination of rational design and 3D-structure-based sequence alignment was applied in the Bornscheuer group and allowed for the design of small focused site-saturation libraries by considering only consensus-like mutations. As a proof of principle these libraries were successfully applied for the development of enzymes with higher thermostability [69] and enantioselectivity [70]. Similarly, also the identification of predicted key residues by rational design methods and applying a subsequent *in silico* alignment search-strategy to identify these residues within existing annotated sequences can lead to the discovery of novel enzymes, as exemplified by the discovery of various R-selective transaminases [71].

In addition, structure-based mutagenesis databases like BRENDA (http://www.brenda-enzymes.org) or the protein mutant database PMD (http://pmd.ddbj.nig.ac.jp), which describe known mutations in enzymes, can be used to identify interesting spots for mutagenesis. Alternatively scanning the whole enzyme sequence by individual replacements to alanine or by site saturation mutagenesis provides information about important functional positions in the protein sequence.

An impressive way to introduce unnatural amino acids based on a combination of (semi-) rational design and directed evolution was presented by Schultz and co-workers [72]. They used structure-guided randomization for the engineering of the substrate binding pocket of aminoacyl-tRNA synthetase to achieve selective binding and charging of orthogonal tRNAs with unnatural amino acids. By alternating rounds of positive and negative selection for the engineered tRNA and against natural tRNAs, respectively, orthogonal unnatural tRNA/aminoacyl-tRNA synthetase pairs were generated. These can now be used to incorporate more than 30 unnatural amino acids in *E. coli*, yeast, or mammalian cells *in vivo* [72].

For some important chemical reactions in organic synthesis no enzymes have been identified yet. Therefore, also computer-based methods have recently become important tools for successful *de novo* design of enzymes. In the Baker lab computational design of catalytic enzyme functions [73] and specificity [74] for challenging chemical reactions have been successfully shown. With a combination of basic molecular orbital theory and protein structure prediction they were able to design a “Diels-Alderase”. Besides this *de novo* design of an enzyme for enantio- and diastereoselective intermolecular reactions [73] they also have shown the possibility to design enzymes for multistep reactions by generating retro-aldol enzymes [75]. The computational design of enzymes offers a multitude of options for basic and applied science, but it also offers much room for improvement. The developed methods are still not reliable enough due to computational complexity and a limited understanding of the subtle amino acid interactions within an enzyme [76].

2.4. Overcoming Limitations of Screening

Finding an improved protein variant relies on smart library design to obtain a high fraction of expressed and functional variants and a statistically relevant screening coverage of libraries. This can be achieved by optimizing the design of libraries based on sequence-based scoring functions, statistical methods, computational enzyme design or combinatorial methods. Typically, manageable library sizes for microplate-based screenings range from 10^3 to 10^5. Nevertheless libraries of up to 10^9 individual variants can be screened employing robotized ultra high throughput screening platforms.

In vivo and *in vitro* display methods have been established to efficiently and reliably screen large protein and peptide libraries, especially for the engineering of protein affinities and stabilities. Display methods, in combination with activity-based fluorescence-activated cell sorting (FACS), can provide an excellent tool for ultra high-throughput screening. The basic principle behind display methods is a physical linkage between genotype and phenotype, either at the cellular level (phage or surface display) or the molecular level of ribosomes or mRNA (ribosome display, ribosome-RNA fusion or *in vitro* compartmentalization by emulsion systems. For enzyme libraries, however, display methods and FACS can only be used if the product(s) of enzymatic conversion or the activity itself can be "linked" to the corresponding displayed enzyme species. The easiest solutions for cell-based systems are fluorogenic substrates which, upon uptake, are intracellularly converted to a fluorescent product that is retained within the cell. This method was successfully applied for the screening of engineered glycosyl transferases in *E. coli* [77]. A summary of other available methods and strategies to provide enzyme –product/activity linkages for FACS-based ultra high-throughput screening is given by Becker *et al.* [78]. Basically, such a linkage can be provided either by covalent attachment of the products to the displayed enzyme or the cell surface, by coupled fluorescent reporter proteins or also by entirely *in vitro* approaches based on emulsion technology.

As such, *in vitro* expression in emulsion is an efficient method to generate large libraries of proteins that are linked to their encoding genes via direct covalent attachment or via beads. However, there is still the problem to pick out the desired protein variants from these large libraries. Dealing with that issue, the groups of Tawfik and Griffiths have shown that compartmentalized *in vitro* [79] and *in vivo* (cell-based) [80] protein expression can be combined with automated screening by fluorescence-activated sorting. This novel combination of methods, termed *in vitro* compartmentalization (IVC), relies on the formation of ~10^{10} single gene (or cell) containing droplets made from water-in-oil-in-water (w/o/w) double emulsions that are amenable to sorting with standard FACS machines.

2.5. Conclusions and Summary

The sheer endless diversity of enzymes in nature has attracted strong notice of biotechnology. A major interest is the application of novel enzymes with new or improved activities in industrial processes as they can contribute to "greener" and more efficient chemical processes. Culture-independent methods such as metagenomics and metatranscriptomics have shown high potential for the successful discovery of new enzymes. However, the exploitation of metagenomes with regard to novel enzymatic activities is still largely limited, mainly because of missing innovative

screening methods. Consequently, to access a bigger fraction of nature's functional diversity, a major focus should be put on developing such screening methods. Although it is difficult to generate novel enzymes *de novo* by laboratory- or computer-based methods recent successful examples showed that this may be an alternative to the discovery of novel enzymes. Additionally, these methods provide excellent tools to adapt and improve existing enzymes to desired properties. The often observed ambiguity of protein structure and protein sequence and the continuously increasing number of available 3-dimensional protein structures makes structure-guided (semi-) rational design, combined with the design of "smart" libraries, a more efficient and powerful method for protein engineering than purely sequence homology-based methods.

Many technology providing companies that were founded during the past 10-15 years drive enzyme discovery and engineering to an optimized and professional level. Nevertheless, new technologies are still speeding up these processes and increase the functional diversity of industrially useful enzymes, thereby promising a strong bio-based industry for the 21st century.

3. Xylanase Discovery by cDNA Library Activity Screening in *S. pombe*

Xylanases are dragging increasing attention due to their wide spectrum of application and the abundance of xylan in nature. In this study, we discovered three novel fungal xylanase sequences. For these purposes, basidiomycota Fomes fomentarius was cultivated on xylan-rich wood and used for RNA isolation. The resulting transcriptome library was expressed in Schizosaccharomyces pombe, and screened for xylanolytic activity. By sequencing the active clones, two different xylanase sequences were obtained, and a third one differing from one of them by a single amino acid. All three were designated as extracellular endo-β-1,4-xylanases belonging to GH family 10; and their closest known homologues showed 74% and 77% identity. The expressed enzymes were characterized, revealing temperature optima between 50 and 60°C, and pH optima between 4 and 5.

3.1. Introduction

Xylan is the second-most abundant polysaccharide in nature, accounting for approximately one third of all renewable organic carbon on Earth [81] . Xylan's abundance arises from it being the main component of hemicellulose - one of the major structural constituents of plant cell walls, together with cellulose and lignin. By its chemical structure, xylan is a heteropolysaccharide formed by a linear backbone of β-1,4 linked D-xylopyranoside monomers, partially substituted with acetyl groups, or sugar residues, such as 4-O-methyl-α-D-glucuronosyl and α-L-arabinofuranosyl. Exact positions and abundance of these substitutions vary between xylans from different sources [82], [83]. Due to the complex structure of xylane, its complete enzymatic degradation requires synergistic action of several enzymes: endo-β-1,4-D-xylanase (EC 3.2.1.8), exo-β-1,4-D-xylosidase (EC 3.2.1.37), α-D-glucuronidase (EC 3.2.1.139), α-L-arabino-furanosidase (EC 3.2.1.55), acetyl xylan esterase (EC 3.1.1.72), ferulic and *p*-coumaric acid esterases (EC 3.1.1.73) [84]. Among these xylanolytic enzymes, endo-β-1,4-D-xylanases (often termed as endoxylanases, or just xylanases [85]) play a major role, as they cleave the β-1,4-glycosidic bonds within the xylan backbone in the initial phase of the breakdown [86], [87]. Hydrolysis by endoxylanases results in xylooligosaccharides, which in turn are suitable substrates for other xylanolytic enzymes [88]. Based on amino acid sequence alignment and hydrophobic cluster analysis, the majority of xylanases have been classified as members of glycosyl hydrolase (GH) families 10 and 11 [89]. Endoxylanases from GH family 10 typically have a higher molecular weight (>30kDa) and a low pI, in contrast to GH family 11 endoxylanases which have a lower molecular mass (<30kDa) and a high pI [90], [91], [92]. Except for GH10 and GH11, xylanases have also been associated with GH families 5, 7, 8, 30, and 43 [81], [93]. Additionally, also bifunctional xylanases/endoglucanases exist in nature [94], [95].

The great focus on xylanase discovery comes from their potential application in a broad range of fields, and the popular idea of using biotechnological alternatives for conventional industrial processes [96]. For example, applying xylanases for bleaching of cellulose pulp in paper manufacturing minimizes the use of chlorine [97]. They are also used in baking industry [98], in textile industry, in waste-paper recycling, animal feed production, bioconversion of xylan to higher value products, bioethanol production, etc. [99], [100].

For these purposes, xylanolytic enzymes have been extracted from a wide range of organisms, including plants and animals [101], [102], but an exceptionally vast reservoir of these enzymes is found in microorganisms [103]. Microorganisms expressing xylanases are found living anywhere from Antarctic [104] to hydrothermal vents of Pacific [105], either like plant pathogens [106] or animal symbioths [107], or free-living organisms. For some bacterial species even ultra large xylanolytic complexes, xylanosomes, have been described [108]. Other valuable source of xylanolytic enzymes are filamentous fungi, which are said to possess the greatest potential as xylanase producers [45]. One of them, basidiomycota *Fomes fomentarius* is a commonly encountered pathogen or saprotroph of beech (*Fagus sylvatica*) and birch (*Betula pendula*) trees in deciduous forests of Central Europe [109]. Like in other angiosperms, xylan accounts for 15 to 30% of the total dry weight in these trees [83]. In order to exploit it, *Fomes fomentarius* must be a highly potent source of strongly expressed and/or very active xylanolytic enzymes. Therefore we saw it as a promising target of search for novel xylanases.

Except for the possible sources, also ways to obtain novel xylanase sequences from nature available for technological applications vary greatly. Diverse strategies to attain novel enzymes have been described in a recent review by [110]. During the 1990s, novel xylanases have mostly been identified by purification directly from homolougos host culture supernatant [111], [112], [113]. In order to gain higher protein amounts convenient for enzymatic characterization, this approach later got combined with heterologous expression in *E. coli* [114], [104] or *Pichia pastoris* [115]. An unusual approach to identify xylanase encoding genes by targeted gene deletion was recently published by [116]. Currently, the most present strategies for acquiring novel xylanolytic enzymes include degenerate PCR using genomic (gDNA), or environmental DNA (eDNA) as templates [117], [105], [118], [93]. Also non-degenerate primers were successfully used in this context [119], but it is evident that the less degenerate a primes is, the smaller are the chances of finding surprising and diverse sequences. Circumventing the involvement of known sequences in enzyme discovery is feasible by applying activity screening. Both genomic and metagenomic expression libraries have been used for this approach. Brennan *et al.* discovered two xylanases with unusual primary sequences and unknown domains by using activity screening of insect-gut metagenome expressed in *E. coli* [120]. Similarly, Mo *et al.* discovered a novel xylanase

by functional screening of a metagenomic library from a rice-straw degrading enrichment culture, using *E. coli* as the expression host [121]. Chanjuan *et al.* and Ko *et al.* used activity screening of *E. coli*-hosted expression libraries in order to find xylanases from bacterial genomic DNA, thereby discovering a novel alkali-stable xylanase [122] and a novel thermostable xylanase [123].

While metagenomics might offer the widest set of available genes, using cDNA as a search field offers much greater focus. Schlacher *et al.* identified a novel xylanase gene from *Thermomyces lanuginosus* with xylanase antibodies used to screen a cDNA phage expression library [124]. A fungal transcriptome library generated from a culture grown on xylan-rich media was used in the work of Luo *et al.* [125] to access a novel GH30 xylanase by random clone sequencing and bioinformatic analysis of the expressed sequence tags. In the past, attempting fungal-enzyme discovery by functional screening of expression libraries has most commonly been done using *Saccharomyces cerevisiae* as the heterologous host [126]. By using this approach numerous fungal plant-polymer degrading enzymes were discovered [127], [128], [129], [130]. Still, *S. cerevisiae* might not be an optimal host for expression libraries, due to it's low expression levels [131]. A possible host overcoming this particular bottleneck is *E. coli*. Its overall popularity as an expression system arises from well characterized genetics and physiology as well as easy and cheap handling [132], [133]. In addition, new techniques allow to circumvent most of the disadvantages commonly connected with *E. coli* expression platform [134], [135]. Still, when it comes to eukaryotic cDNA libraries and attempting to express a large number of unknown proteins, there is no room for careful optimization of expression. Instead, a robust and reliable eukaryotic expression system is needed.

Various yeasts are industrially used as expression hosts for enzyme production. However, fission yeast *Schizosaccharomyces pombe* is attracting a lot of attention due to various important properties that it shares with higher eukaryotes [136]. It was the third eukaryotic organism to be fully sequenced, right after *Saccharomyces cerevisiae* and *Homo sapiens*, and its value as a eukaryotic model-organism is indisputable. Nevertheless, despite a lot of work that has been done on engineering it for more efficient heterologous expression [137], *S. pombe* is still far less established in industry then other yeasts [138]. Since high protein yields are not essential for library activity screening, but rather proper folding and functional expression, *S. pombe* could have the advantage over other more established systems [139]. Within the work of Gruber and Abad [140] *S. pombe* indeed confirmed to be a promising host for expressing fungal gDNA libraries. One bottleneck of this approach was the size of commercially available vectors and a low transformation efficiency, which in the case of libraries can mean a detrimental loss of valuable sequences. In our work we used an *E. coli*– *S. pombe* shuttle vector (pGAZ2) specially designed [141] to overcome this limitation. The pGAZ2 vector is significantly smaller then any other available *S. pombe* plasmid, and

therefore able to carry large inserts. Its small size is partially due to having only a single antibiotic-resistance cassette for both *E. coli* and *S. pombe*. The antibiotic resistance also accounts for the advantage of using wild type *S. pombe* and rich media, instead of slow-growing auxotrophic strains and complicated minimal media connected with the thiamine repressed promoters of commercial vectors. pGAZ2 uses a constitutive yeast promoter to drive heterologous gene expression.

In this study, we describe a attempt to discover novel xylanase sequences by activity screening of *F. fomentarius* transcriptome library, using an unconventional heterologous expression system *S. pombe* and a custom-designed plasmid. This approach yielded three novel xylanase sequences. Although lignocellulose degrading enzymes have been isolated from naturally grown *F. fomentarius* before [142], these xylanases are the only characterized recombinant proteins from *F. fomentarius* so far.

3.2. Materials and Methods

3.2.1. *F. fomentarius* cultivation, RNA isolation and Library Construction

F. fomentarius strain (DSM-3520) was grown on agar containing peptone, with an addition of 2% w/v grinded beech wood. Incubation was done at 30°C for 12 days. For RNA isolation 700mg of *F. fomentarius* mycelium was scraped off the plate and grinded down to homogenous powder using liquid nitrogen, mortar and a pestle. Total RNA isolation was done using Ambion RNAqueous kit (Life Technologies, Paisley, UK). Reverse transcription, normalization, and library construction was done by LGC Genomics (Berlin, Germany). The library was directly cloned in *E. coli-S. pombe* shuttle vector pGAZ2. For this purpose, a part of the pDNR-LIB vector sequence was introduced in pGAZ2, containing required *SfiI* cloning sites. Precise modification of the pGAZ2 vector, as well as the cloning strategy, are explained in chapter 5.2 (**Figure 16/17**).

3.2.2. *S. pombe* Transformation

Protocols for manipulating *S. pombe* cells were adapted from [143] and [144]. For preparing competent cells, a *S. pombe* preculture was made by inoculating 50ml yeast-extract media containing 100mg Zeocin/l and 1.25g/l of *S. pombe* amino acid supplements (Formedium, Hunstanton, UK) /l (YES-Zeo), and cultivated at 30°C over night. The preculture was used to inoculate 500ml of YES-Zeo and create the main culture. The main culture was grown to cell density of 10^7 cells/ml, and before harvesting it was placed on ice for 15min. Harvesting was done by centrifugation for 5min at 1600g. The resulting pellet was resuspended in 50ml DTT buffer (20mM hepes, 0.6M sorbitol, 25mM DTT) and incubated at 30°C for 15min. The cells were centrifuged

again (5min, 1600g), and washed three times using 50ml of ice-cold 1M sorbitol. Finally, the pellet was resuspended in 60ml 2M sorbitol, and 300µl aliquots were placed at -80°C to freeze gradually, and stored there for future transformations. For transformation, aliquots were thawed at 30°C and then centrifuged. The resulting pellet was washed once with ice-cold 1M sorbitol, and finally resuspended in 40µl 1M sorbitol. The cells were transferred in a pre-chilled cuvette (0.2cm electrode gap) and mixed with 100ng DNA. Electric pulse of 2kV was applied, and 0.5ml of ice-cold 1M sorbitol was added immediately. Transformed cells were kept at 4°C for 90min, and after adding 0.5ml YES media they were incubated at 30°C for another 90min, with shaking. Transformations were plated on YES-Zeo agar plates, which were then incubated at 28°C for three days.

3.2.3. *S. pombe* Picking and Activity Screening

S. pombe colonies were picked to 384 well plates (50µl YES-Zeo/well) using the picking robot (QPix2, Genetix). The 384 well plates were incubated for 2 days at 28°C, in a glass box saturated with humidity to prevent evaporation of the media. After this, 30µl 50% glycerol was added to each well, gently shaken for 30 min, and placed at -80°C. These glycerol stocks were then used to inoculate deep-well plates (DWP) filled with 500µl YES-Zeo media/well. The DWP carrying single *S. pombe* clones from the library were incubated at 30°C, 300rpm, for three days. Supernatants were collected by 10min centrifugation at 3200rpm. 50µl of the supernatant were mixed with 50µl of the substrate mix, and the sealed black micro-titer plates with the reaction mix were incubated for 6h at 37°C with shaking. The substrate mix contained 4-methylumbelliferyl-β-D-xylopyranoside (MUX) in 10µM concentration, and 5µg/ml of 4-methylumbelliferyl-β-D-xylobioside (MUX_2) (both dissolved in 25mM Na-acetate buffer, pH 4.8). After the incubation, the reaction was quenched by adding 100µl 1M Na_2CO_3, and fluorescence was measured (λ_{exc}=365nm, λ_{em}=450nm). As a negative control *S. pombe* strain carrying empty pGAZ2 vector was used. Rescreening with various methylumbelliferyl (MU) carrying substrates was done in the same way like screening, except that 50µl of 100µM (50µg/ml in case of MUX_2) single substrate solutions were used for the assay (all dissolved in 25mM Na-acetate buffer, pH 4.8). Following substrates were tested in rescreening: 4-methylumbelliferyl-α-L-arabinofuranoside (MUAf), 4-MU-β-D-cellobioside (MUC), 4-MU-α-D-galactopyranoside (MUGal), 4-MU-β-D-glucopyranoside (MUGlc), 4-MU-β-D-glucuronide hydrate (MUGh), 4-MU-β-D-lactopyranoside (MUL), 4-MU-β-D-mannopyranoside (MUM), 4-MU-α-L-rhamnopyranoside (MUR), MUX, and MUX_2. All fluorescent MU-substrates were ordered from Sigma-Aldrich (St. Louis, MO), except for MUX_2, which was ordered from Life Technologies (Paisley, UK) as a part of the EnzCheck© *Ultra* Xylanase Asay Kit.

3.2.4. Sequencing and PCRs

The detected active clones were sequenced using outer multiple cloning site primers for the pGAZ2 vector. All primers were ordered from Integrated DNA Technologies (IDT) (*Coralville IA, USA).* The amplifications were done using the Phusion polymerase (Finnzymes, Espoo, Finland). The PCR conditions were set according to manufacturer's manual, taking into account the length of the amplified fragment, and primer melting temperatures. All products obtained from PCRs were purified with the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany).

3.2.5. *S. pombe* Fermentation

S. pombe clones 5I9, 8N23, and 9P23, as well as the negative control strain (*S. pombe* carrying empty pGAZ2) were cultivated in 5l BIOSTAT C plus bioreactors from Satorius (Göttingen, Germany). The initial preculture was made by inoculating a single colony into 3ml YES-Zeo media, and grown for two days at 28°C. This culture was then used to inoculate 50ml YES-Zeo. After another two days of growing at 28°C, the 50ml culture was added to 250ml YES-Zeo media, in order to get 300ml of the preculture. This final preculture was grown over night at 28°C, until reaching OD_{600} 13-16. In the bioreactors 3l YES-Zeo media was inoculated by 300ml preculture, giving a starting OD_{600} of about 1.5. Temperature was maintained at 30°C, and pH at 5.0. Oxygen saturation was regulated by stirring and airflow, and pO_2 was kept above 30%. Glucose feed was started after pO_2 initially started rising, which was approximately around 20h from the cultivation beginning. Till the end of cultivations, around 500g of 60% glucose was added per bioreactor. The cultivations were harvested after 63 to 73h.

Samples were taken regularly for each cultivation, and OD_{600} was measured immediately to monitor the cell growth. The rest of the sample was centrifuged, and resulting supernatants were stored in two aliquots, one at +4°C, and the other at -20°C. The supernatant samples were later analyzed for activity in the bellow mentioned way.

Harvested supernatants were concentrated using cross-flow filtration (Quattroflow, Kamp-Lintfort, Germany) with 10kDa cutoff. Starting volume of >3000ml was reduced to a volume of about 150ml. Concentrates were aliquoted and stored at -20°C. Activity was measured for the concentrated supernatants, the filtration flow-trough, as well as for all the samples collected throughout the fermentation. The assay was done using 10µl 4-methylumbelliferyl-β-D-xylobioside (MUX_2) (50µg/ml), 10µl supernatant, and 80µl 50mM Na-acetate buffer, pH 4.8. Reactions were incubated with shaking at 37°C, for one hour. Quenching and measuring was done as described in chapter 3.2.3.

3.2.6. Assay Linearity and Enzyme Characterization

To assure a linear range of the assay, we prepared following dilutions of the supernatant: 1:1, 1:2, 1:5, 1:10, 1:20, 1:50, 1:100, 1:200, 1:500, and 1:1000. For the assay 20µl of diluted supernatant was mixed with 10µl substrate (MUX_2, 50µg/ml), and 70µl buffer (25mM sodium acetate buffer, pH 4.8). Reaction incubation was done at 37°C, for 45min, with 300rpm shaking. The reaction was quenched, and fluorescence measured as described in chapter 3.2.3. All measurements were done in triplicates, and if not stated otherwise, all further characterization assays were done in this manor.

For determining pH profiles of selected enzymes, reactions were set including different buffers, pH ranging from 3.5 to 8.0. For pH range between 3.5 and 5.5 we used 0.1M sodium acetate buffer, and for range from 6 to 8 we used 0.1M sodium phosphate buffer. 20µl supernatant was mixed with 10µl substrate, and 70µl of appropriate buffer. In order to determine the temperature profile of selected enzymes, a reaction was set in the following way: 40µl supernatant was mixed with 20µl substrate, and 140µl buffer. Incubation was done at temperatures ranging from 10°C to 80°C, with 10° steps. After incubation, 100µl of reaction mixture was transferred to a black microtiter plate, then quenched and measured as mentioned above. The temperature stability was determined by doing an assay using supernatants previously incubated for one hour at different temperatures. This pre-incubation was done at 4°C, and at temperatures ranging from 10°C to 90°C with 10°C steps.

3.3. Results

3.3.1. Library Screening and Sequence Analysis

About 22000 *S. pombe* clones were screened using the described fluorescence based xylanase assay including MUX and MUX_2 as susbtrates. As a criterion for a clone to enter the rescreening round we chose a low cut-off of at least two fold fluorescence intensity of the negative control. By these criteria, 19 clones were selected for rescreening, which further disclosed 5 active clones after rescreening. Sequencing of these 5 clones revealed 3 unique sequences: an identical sequence was found for clones 8N23 and 9F22, while clone 5I9 differs from them by two silent mutations and one amino acid substitution; further also clones 9P23 and 10J24 have shown to have an identical sequence. The nucleotide sequences are represented in **Figure 5**, while encoding protein sequences and their alignment are shown in **Figure 4**. Clones 5I9, 8N23, and 9F22 have protein sequence length of 351 amino acid, while sequences of clones 9P23 and 10J24 are 394 amino acids long. Sequence analysis confirmed the

presence of a secretion signal peptide for all 3 sequences, and according to NCBI's blastx homology searches all 3 were found to be members of the GH10 family, most probably endo-1,4-β-xylanases. Xylanases of clones 9P23 and 10J24 were accounted with a fungal carbohydrate-binding module (CBM) as well. The closest known homolog of xylanases from clones 5I9, 8N23, and 9P23 is a GH10 family protein from *Dichomitus squalens*, a basidiomycota belonging to the same family of bracket-fungus (*Polyporaceae*) as *Fomes fomentarius*. These proteins exhibit 74% identity of the primary sequence. Xylanases from clones 9P23 and 10J24 show highest identity, namely 77%, to another GH10 family member coming from the above mentioned organism, *Dichomitus squalens*. Clones 5I9, 8N32, and 9P23, representing three different sequences found, were chosen for rescreening on a wider set of substrates and for further analysis.

Testing of these clones on additional MU substrates, revealed most prominent activity towards endo-β-xylanase substrate MUX_2, which was in accordance with the activities suggested by homology searches. The clones 5I9 and 9P23 showed additional activity towards the exo-β-xylosidase substrate, MUX, as well as towards MUC, the cellulase substrate (**Figure 6**), while the 8N23 clone exhibited only endoxylanase activity. This was an interesting finding, considering that it differs from clone 5I9 by a single amino acid.

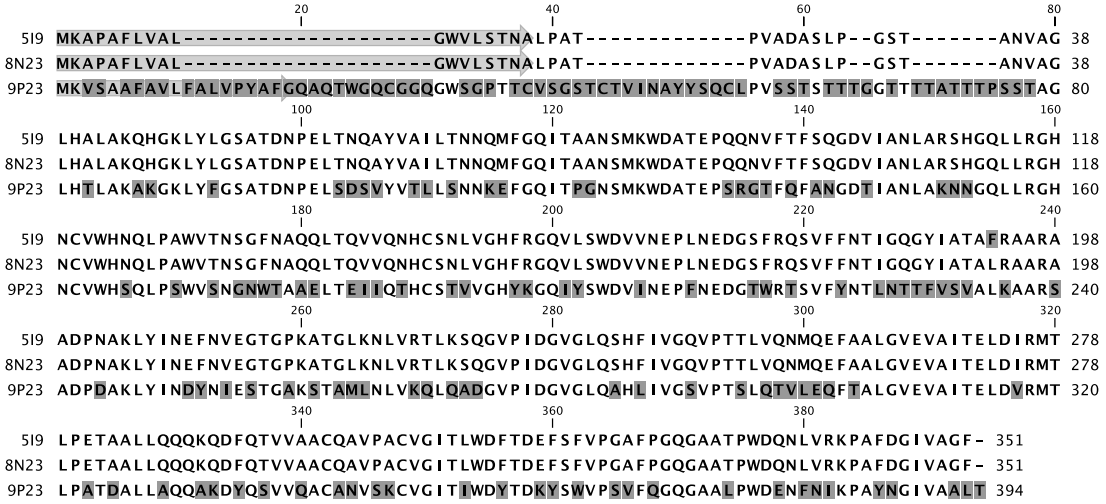

Figure 4 – Protein sequences obtained by translating nucleotide sequences of the 5 active clones emerging from *F. fomentarius* library screening, and their alignment with each other. Sequence of clone 5I9 is unique; sequence of clone 8N23 is identical to clone 9F22 (not shown); sequence of clone 9P23 is identical to clone 10J24 (not shown). Residues differing between the clones are highlighted in dark gray. Clone 5I9 differs from clone 8N23 by a single amino acid at position 193. The predicted secretion signal peptides are indicated by light gray arrows.

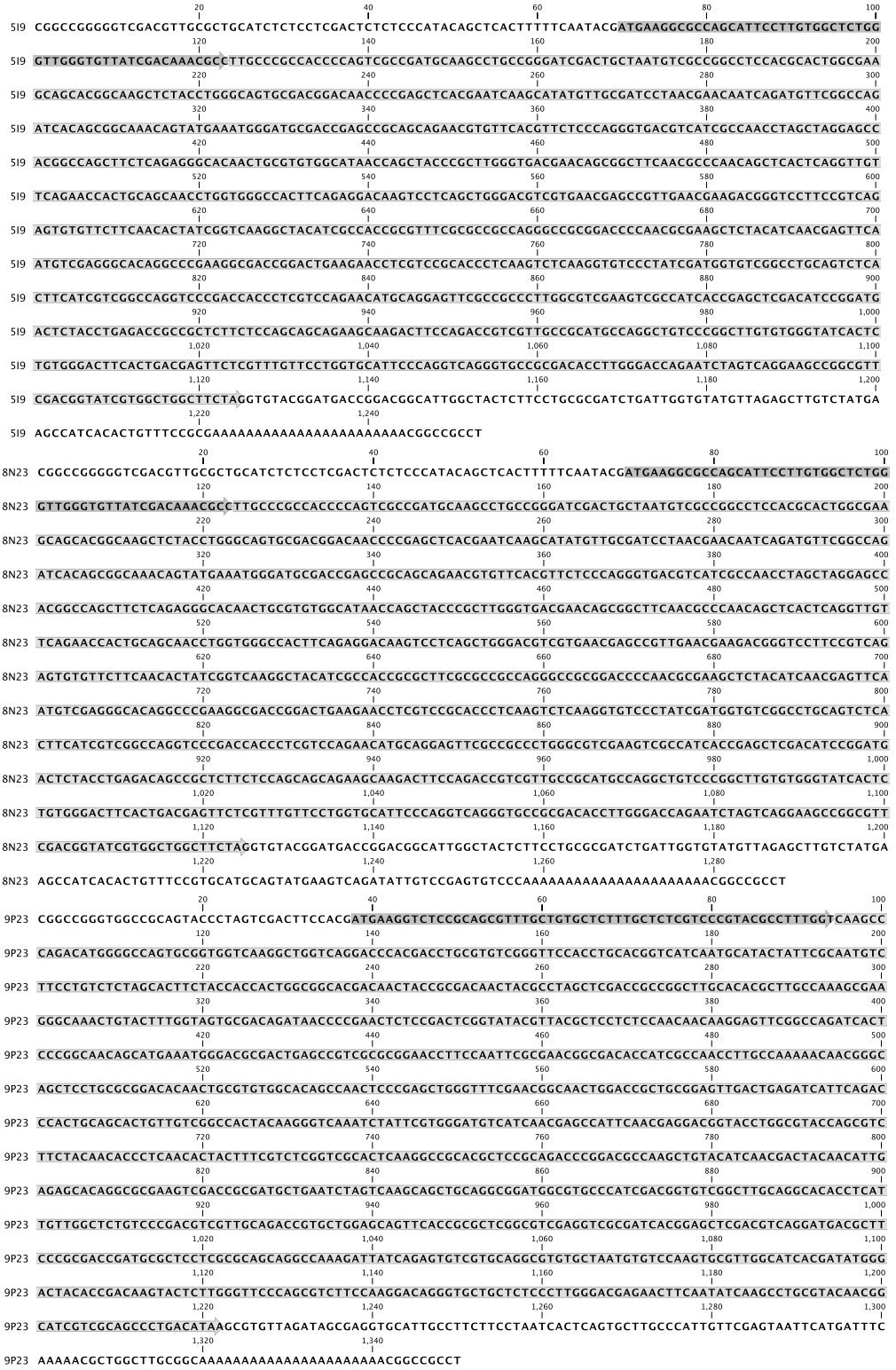

Figure 5 – Nucleotide sequences obtained by sequencing of the 5 active clones emerging from *F. fomentarius* library screening. Sequence of clone 5I9 is unique; sequence of clone 8N23 is identical to clone 9F22 (not shown); sequence of clone 9P23 is identical to clone 10J24 (not shown). Signal peptides are indicated with a dark grey arrow, while a lighter grey arrow indicates the rest of the coding sequence.

Namely leucine at position 193 in clone 8N23 is exchanged to phenylalanine in clone 5I9 (L193F, **Figure 4**). It remains to be shown, if this sequence difference is due to the existence of different isoenzymes in nature or due to an error introduced during cDNA library construction or reverse transcription. In general, it is not surprising for GH10 xylanases to show activity towards xylosidase substrates, and their relatively wide substrate binding cleft is known to be tolerant to the replacement of one or two consecutive D-xylopyranosyl residues by D-glucopyranosyl residues in substrates [145], [146].

3.3.2. *S. pombe* Cultivation

Cultivation of the selected clones in the bioreactors was not optimal, yielding cell densities OD_{600} of maximum 16. Still, further attempts to optimize the fermentation process have not been made, as the yielding protein amounts were sufficient for enzyme characterization (data not shown).

3.3.3. Enzyme Characterization

Taking into account the linear range of activity in dependence of enzyme concentration, following dilutions were chosen for further characterization: 1:60 for clone 5I9; 1:20 for clones 8N23 and 9P23. Using this assay set up, pH profiles were determined, and shown in **Figure 7**. Within the pH range from 3.5 to 8.0, the tested enzymes showed highest activities in slightly acidic environment. Xylanase from clone 9P23 exhibited its maximal activity at pH 5, while being highly active also between pH 4.0 and 5.5. Xylanase from clone 5I9 showed an almost identical pH profile, with the optimum pH being 5. Xylanase from clone 8N23, although differing from 5I9's xylanase by a single residue, showed the peak of its activity at pH 4.

Also temperature profiles were determined for the three clones, and are represented in the **Figure 8**. Xylanases from clones 5I9 and 9P23 showed an activity peak at 60°C, while xylanase from clone 8N23 once again stood out, with a temperature optimum being at 50°C. Despite the minimal difference in primary structure between this enzyme and the xylanase from clone 5I9, it showed lower temperature stability as well (**Figure 9**): 8N23-xylanase retained only 60% of it's maximal activity already after one hour pre-incubation at 50°C. Other two xylanases, from clones 5I9 and 9P23 were able to withstand one hour pre-incubation at 60°C, and still retain more than 90 and 85% of their maximal activity (respectively).

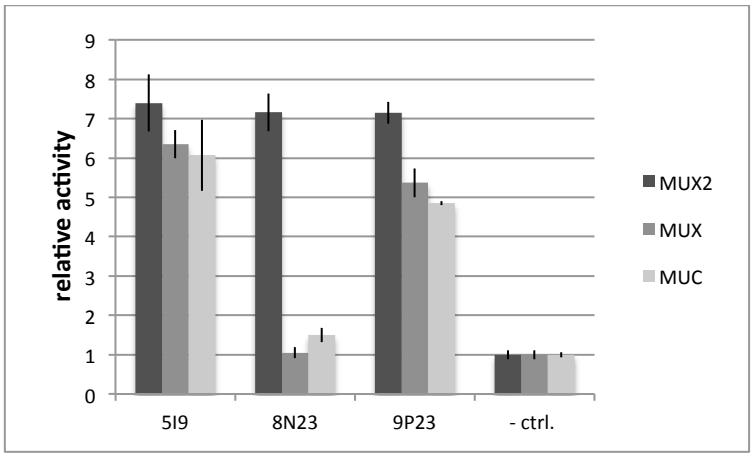

Figure 6 – Activities of clones 5I9, 8N23, and 9P23 relative to the negative control (-ctrl). Clones 5I9 and 8N23 differ by a single amino acid, yet they show significant difference in activity pattern. MUX_2 is an endo-β-xylanase substrate, MUX is an exo-β-xylosidase substrate, while MUC is a typical cellulase substrate. Activity towards MUX_2 is likely to be even much higher, but the substrate concentration has most probably been limiting in this initial trial assay.

3.4. Discussion

Starting off the search for novel xylanases, we aimed at *F. fomentarius* as a potential source. Theoretically, *F. fomentarius* is bound to be a rich source of xylanolytic enzymes, as it feeds on beech and birch wood, both containing significant share of xylan (15-30%) [83]. In addition, *F. fomentarius* has experimentally proven its xylanolytic activity within the work of *Větrovský et al.* Moreover, they observed an effect of wood type on the enzyme production: the activity of xylan-degrading enzymes was significantly higher in beech wood than in birch wood [109]. When cultivating our source organism, we made sure the xylanase-encoding genes are active by growing it on inductive media, containing beech wood. Therewith, isolating total RNA yielded a snapshot of transcribed genes, including those encoding for xylanolytic enzymes. In order to detect those genes, we relied on *S. pombe* to correctly express the cDNA coming from *F. fomentarius*. We used a fluorogenic endoxylanase substrate to screen about 22000 clones for traces of activity. Considering that sequencing of the same *F. fomentarius* transcriptome (done within a separate approach, data not published) reported about 8000 contigs bigger then 500bp; screening of 22000 clones provided an approximate 2.8 fold coverage of the library. Most accurately, sufficient library coverage can be evaluated by independently discovering multiple positive clones carrying identical DNA sequences. In our study, we indeed detected two identical clones active on the endoxylanase substrate, namely 9P23 and 10J24, and further also identical clones 8N23 and 9F22. The fifth clone, 5I9, turned out to contain

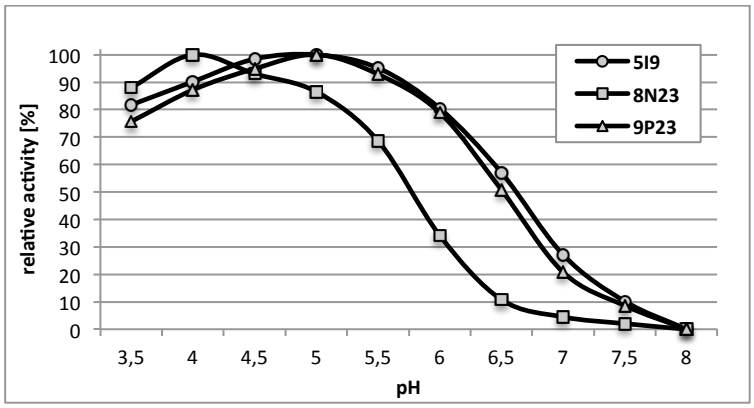

Figure 7 – Comparison of pH profiles of expressed xylanases of 3 selected *S. pombe* clones. Relative activities are expressed in percentages, where 100 represents individual highest activity for each clone, and 0 represents individual lowest activity for each clone. A similar profile can be observed for all 3 clones, with pH optima ranging from 4 to 5.

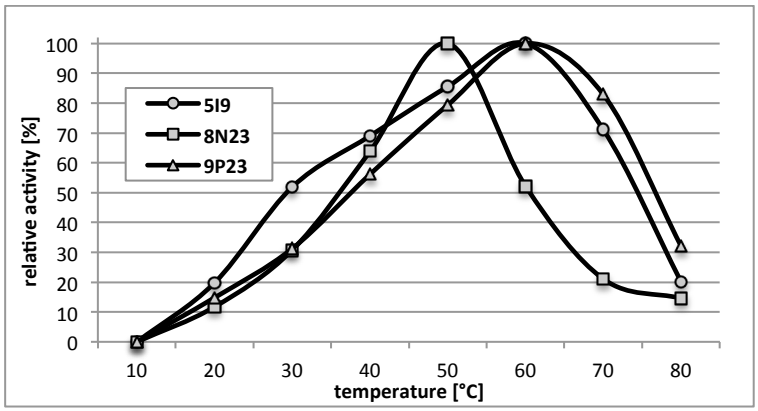

Figure 8 – Temperature profile comparison between expressed xylanases of 3 selected *S. pombe* clones. Relative activities are expressed in percentages, where 100 represents individual highest activity for each clone, and 0 represents individual lowest activity for each clone. Clone 8N23 shows optimal activity at 50°C. Clones 5I9 and 9P23 show a bit higher temperature optimum, at 60°C.

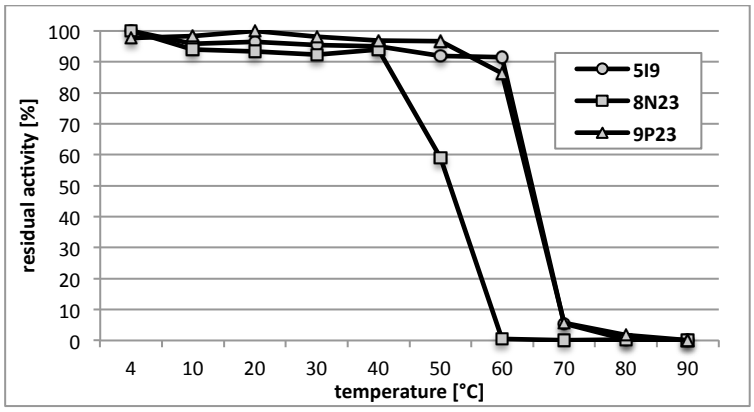

Figure 9 – Thermal stability comparison between expressed xylanases of 3 selected *S. pombe* clones. Relative activities are expressed in percentages, where 100 represents individual highest activity for each clone, and 0 represents individual lowest activity for each clone.

a variation of the sequence of the latter two clones, differing by two silent, and one point mutation. In comparison with clone 8N23, the single amino acid exchange (L193F) accounted for additional activity of clone 5I9 towards exo-β-xylosidase substrate (MUX) and cellulase substrate (MUC). Moreover, clone 5I9 showed different temperature properties in comparison to clone 8N23. Higher temperature optimum and higher stability, as well as higher pH optimum of 5I9-xylanase must be accounted to above mentioned amino acid exchange, giving it even more significance. According to sequence analysis and homology searches, all three identified sequences were designated as extracellular endo-β-1,4-xylanases belonging to GH family 10. The closest known homologues for xylanases expressed by clones 5I9/8N23 and 9P23 showed 74% and 77% identity (respectively). The enzyme characterization in respect to pH and temperature preferences showed no abnormalities compared to an average fungal xylanase [45, 88, 96]. The pH optima are lying within the acidic range, 4 for clone 8N23 and 5 for clones 5I9 and 9P23; while the optimal temperature is 50°C for clone 8N23, and 60°C for clones 5I9 and 9P23.

Although this study didn't yield xylanases of extraordinary properties, it is a valuable example of exploiting *S. pombe* for a completely new purpose, as a host for eukaryotic expression libraries. With its inborn advantages based on similarities to higher eukaryotes, and a custom-made expression vector, *S. pombe* can now be considered a reliable high-throughput expression system, and worthy addition to existing pool of established heterologous hosts [147]. The discovered xylanases themselves open new ground for protein engineering towards more industry-desirable properties, with an initial hint on position that might contribute to both substrate specificity and enzyme stability.

4. Thermostability Improvement of Endoglucanase Cel7B from *Hypocrea pseudokoningii*[*]

Exploiting enzymes for industrial purposes often requires engineering of these enzymes to adapt them to the industrial requirements. In order to meet industrial demands, we improved the thermostability of endoglucanase Cel7B from Hypocrea pseudokoningii (HpCel7B), which was heterologously expressed in the yeast P. pastoris. Random mutants showing higher activity at elevated temperature have been selected and sequenced. In addition a model structure of our target enzyme was compared to structures of homologous but more thermostabile endoglucanases. This comparison pointed out several potential hot spots that were recognized as important for thermostability. The most promising mutations from both rational and non-rational approaches were randomly recombined by gene synthesis to evaluate potential additive effects for thermostability. This recombination library yielded a number of improved variants, of which the best ones were sequenced and characterized. Compared to the starting variant, recombination mutants showed up to 10°C higher melting temperatures and can be used at higher temperature than the natural enzyme.

4.1. Introduction

Lignocellulose, as the most abundant carbohydrate polymer in nature [148], has great potential as a sustainable raw material for biochemical and biofuel production. Despite many advantages, commercial processes for lignocellulose utilisation still seem to be too expensive due to its extreme recalcitrance [1, 149]. However this is currently changing due to new tools and technologies. The biotechnological lignocellulose exploitation via total hydrolysis consists of three stages: (I) physical and chemical pretreatment and fractionation of raw plant biomass, (II) enzymatic degradation of the pretreated fibres, and (III) biotransformation of released sugars. Among these three steps, a key-challenge is the enzymatic degradation of cellulose fibres and the release of fermentable sugars in a cost effective way [150]. Hydrolytic enzymes degrading cellulose and hemicelluloses are required for this step and are supported by redox enzymes and esterases. Cellulose hydrolysing enzymes are major components of efficient enzyme cocktails and they belong to three different families: endoglucanases, which cut randomly at internal amorphous sites in the polysaccharide chains; then cellobiohydrolases, that act at the ends of the chains and release glucose dimers; and finally β-glucosidases, that degrade those dimers to single glucose molecules [19].

In many cases, natural enzymes are poorly suited for industrial applications and often need to be adapted to withstand extreme conditions, such as high temperature [151]. Elevated temperatures of industrial processes allow higher reaction rates, better solubility of reactants, and solve microbial contamination issues. Therefore, the

[*] Mitrovic, K. Flicker, G. Steinkellner, K. Gruber, C. Reisinger, G. Schirrmacher, A. Camattari, A. Glieder; *Journal of Molecular Catalysis B: Enzymatic*, 2013; submitted

thermostability is a highly desired property for enzymes in industrial biorefining [152]. Contributing to the global efforts towards optimizations in enzymatic cellulose degradation, we have been engineering the thermostability of an endoglucanase Cel7B originating from *Hypocrea pseudokoningii* (*Hp*Cel7B). *Hp*Cel7B belongs to glycosyl hydrolase family 7 (GH7), and it is a single module enzyme. Its overall fold consists of two anti-parallel β-sheets forming a β-sandwich. In addition to GH7, this fold is also characteristic for family 12 enzymes.

To achieve functional expression of *Hp*Cel7B, we opted for *Pichia pastoris* as a host. Ability to perform necessary posttranslational modifications such as N-glycosylation and disulphide bond formation gives *Pichia pastoris* a reputation of a good host for the expression of secreted eukaryotic proteins. Other advantages include high protein yields combined with growing to very high cell densities and low levels of secreted intrinsic proteins [153]. Libraries for protein engineering can be constructed by integrating so called linear expression cassettes (LEC) into the genome [154]. Linear expression cassettes are DNA fragments assembled by PCR, thereby circumventing intermediate cloning and loss of diversity by *E. coli* transformation.

As there is no general distinct structural feature that accounts for higher stability of thermophilic proteins and no general strategy for thermal stabilization [155], [156], random mutagenesis and directed evolution have often been used for stability improvement of various enzymes (for a recent review see [157]), including different endoglucanases. For example *Liang et al*. used directed evolution to broaden the temperature profile of an already thermophilic endoglucanase Cel5A [158], *Nakazawa et al.* used it to improve thermo- and pH stability of *T. reesei* endoglucanase III Cel12A [159]. Directed evolution is without doubt a powerful strategy for such challenges. However, some amino acid exchanges require more than a point mutation, and also desired synergistic mutations at different sites are hard to achieve by mere random mutagenesis.

Random mutagenesis can be complemented with different (semi)rational techniques which are systematically overviewed in a recent review by *Lutz* [160]. The advantage of complementing a random approach with rational design is well demonstrated in recent work of *Anbar et al.* who initially improved thermostability of endoglucanase Cel8A from *Clostridium thermocellum* by directed evolution [161], and later further improved it by adding two more single mutations suggested by rational design [162]. *Yi et al.* worked on the thermostability of the same endoglucanase and complemented the previously employed random shuffling approach with knowledge-based rational design [163]. Also *Voutilainen et al.* used rational design to increase the thermostability improvement of endoglucanase Cel7B initially achieved by random mutagenesis [164]. However, rational design to achieve greater protein stability is a challenging task, as sequence and structural analysis cannot give definite predictions

for stability improvement [165]. The literature data is often conflicting, and features that cause thermostability for one protein family are frequently not significant for other families. Moreover, preferred amino acids and intramolecular interactions leading to thermostability differ not only among protein families, but might be specific to certain folds [166]. Nevertheless these indications made by *Yennamalli et al.* are a result of theoretical sequence and structure analysis, and have not been experimentally confirmed.

Still there are a few protein modifications that are repeatedly occurring in literature as beneficial for thermostability: *Voutilainen et al.* used disulphide bridges to improve two different cellobiohydrolases from glycohydrolase family 7. They further increased the thermostability improvement of Cel7B from *Melanocarpus albomyces* by introducing an additional S-S bridge [164]. Similarly, in case of Cel7A from *Talaromyces emersonii* they raised the T_m for 10°C by introducing 3 additional S-S bridges. Probably due to higher rigidity also the activity at temperatures lower than 75°C was decreased [167]. Further, the carbohydrate-moiety has long ago been suggested as a stabilizer of thermophilic *Humicola insolens* endocellulases [168] and also of β-glucanases [169], [170], as well as other proteins [171]. More recent research of *Voutilainen et al.* demonstrated that N-glycans contribute to the stability of the cellobiohydrolase Cel7A from *Talaromyces emersonii* [167]. Since hyperglycosylation of cellulases recombinantly expressed in *S. cerevisiae* might actually decrease their activity [172], [173], [174], [175], [176]; *P. pastoris* appears to be a good alternative expression host when it comes to thermostability engineering. Relaying on experimental findings and theoretical comparisons, another general conclusion can be made: less cavities relate to better stability [177], [178], [179], [180]. In terms of activity conservation, the surface cavities seem to be less risky targets for mutagenesis than the core cavities. The reason for this is that the surface residues are less involved in intramolecular interactions comparing to the core residues, and therefore altering them shows less effect on protein folding [181]. The importance of surface mutations in engineering towards thermostability was demonstrated e.g. by *Martin et al.* [182], and more recently by *Kotzia et al.* who suggested a stabilizing effect of an amino acid exchange towards a more hydrophobic residue on the surface [183].

To achieve our goal to stabilize an endoglucanase, we combined random mutagenesis with (semi)rational design. For the prediction of potentially beneficial mutations, we relied on above-mentioned structural characteristics and employed sequence/structure comparisons of *Hp*Cel7B with more stabile homologues. Although natural sequence diversification often does not influence the properties of homologous enzymes, recruiting substitutions from more thermostabile homologues, or even from family consensus, was previously proven as useful in case of endoglucanase Cel12A from *Trichoderma reesei* [184], [185], and endoglucanase Cel8A from *Clostridium thermocellum* [162].

4.2. Materials and Methods

4.2.1. Expression Strategy

To achieve optimal expression of the wild type *Hp*Cel7B gene (GenBank accession number: ABM90986.1), four different LECs were constructed, combining two different promoters (inducible AOX promoter and constitutive GAP promoter) with two different secretion signals (the native one, and the secretion signal of *Saccharomyces cerevisiae* α-mating factor) and the endoglucanase gene as well as with the selection marker. Overlap extension PCR, as described by *Liu et al* [154], was used to construct the LECs.

4.2.2. Random Mutagenesis and Library Construction

A library of *Hp*Cel7B mutants was produced by error-prone PCR of the endoglucanase gene, using the *Taq* polymerase (Promega, Madison WI, USA), Mg^{2+} concentration elevated to 7mM, and a nucleotide imbalance (dATP, dGTP: 0.2mM each; dCTP, dTTP: 1mM each). All primers (throughout the whole project) were ordered from Integrated DNA Technologies (IDT) (Coralville IA, USA). The PCR conditions were following: 2min at 95°C, 30 cycles of [1min at 95°C, 1min at 56°C, 1min at 72°C], 5min at 72°C. All products obtained from PCRs were purified with the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany).

4.2.3. Linear Expression Cassette (LEC) Construction

LECs were constructed by overlap extension PCR (oePCR) using the Phusion polymerase (Finnzymes, Espoo, Finland), and three overlapping DNA fragments in equimolar amounts: 1) a fragment containing the GAP promoter (or AOX promoter for initial testing of the wild type expression); 2) a fragment containing the pool of randomly mutated *Hp*Cel7B genes (or the wild type *Hp*Cel7B gene for constructing the positive control LEC); 3) a Zeocin resistance cassette. The first step of oePCR was done under the following conditions: 30 sec at 98°C, 10 cycles of (10sec at 98°C, 20sec at 60°C, 60sec at 72°C), and 7min at 72°C. In the second step, after the addition of terminal primers the conditions were: 30sec at 98°C, 30 cycles of (10sec at 98°C, 20sec at 60°C, 80sec at 72°C), and 7min at 72°C. A negative control LEC was prepared by omitting the middle fragment.

4.2.4. Rational Design

A 3D homology model of *Hp*Cel7B, based on the highly similar protein sequence of *Trichoderma reesei* endoglucanase 1 (*Tr*EG1; PDB ID: 1EG1), was generated using YASARA Structure [186]. Cavities were calculated using a LIGSITE algorithm [187]. For the analysis of the hydrophobicity of the cavities, the hydrophobic calculation part of the program VASCo [188] was used. The figures were created using PyMOL (Schrodinger, LLC: The PyMOL Molecular Graphics System, Version 1.3r1). Mutation suggestions were made based on structural and sequence comparison with *Fusarium oxysporum* and *Humincola insolens* endoglucanase 1 (*Fo*EG1, *Hi*EG1; PDB IDs respectively: 3OVW, 2A39), and *Tr*EG1 (PDB ID: 1EG1). The flexible regions of *Hp*Cel7B were predicted based on calculation of B-factors of the *Tr*EG1 protein structure.

4.2.5. Site Directed Mutagenesis

The primers carrying the desired mutations were used to create PCR fragments that were later assembled into complete LECs by overlap extension PCR (as described in 4.2.3). All fragments were amplified by Phusion polymerase using the following programme: 30sec at 98°C, 30 cycles of [10sec at 98°C, 20sec at 60°C, 6-75sec at 72°C - depending on the fragment length and according to the manufacturer's information on enzyme's processivity], and 7min at 72°C.

4.2.6. Recombination Library

The recombination library was provided as a service by Sloning Biotechnology (Munich, Germany), by combining ready-made nucleotide triplets as gene building blocks. For the defined positions to be mutated, a 1:1 mixture of wild type and the desired mutant triplet was used for the extension, thereby reducing useless complexity of the library compared to degenerated oligos.

4.2.7. Library Transformation and Cultivation

Pichia pastoris cells were prepared and transformed based on a protocol by [189]. Transformants were selected on YPD agar plates containing Zeocin 100mg/L, and picked to 384 well plates (YPD media, 50ml/well) by picking robot (QPix2, Genetix). After 3 days of incubation at 28°C in a moisture saturated box, 20µl of 50% glycerol was added to each well. After 20min shaking the plates were stored at -80°C. These plates containing the glycerol stock of the library were used as master plates for inoculating DWPs (BMD5% 250µl/well) by stamping. Each DWP contained also the

reference strain (wild type *Hp*Cel7B) and a negative control strain. Inoculated DWPs were cultivated for 60h at 28°C, 80% humidity, and 280rpm.

4.2.8. Activity Assays

DWP cultivation supernatants were collected after 10min centrifugation at 4°C, 3200g. The MUL (4-methylumbellyferryl β-D-lactopyranoside) activity assay was performed as follows: 10µl of the cultivation supernatant were mixed with 90µl 100µM MUL in 25mM Na-acetate buffer with pH 4.8. Plates were sealed and incubated for 2h, with 300rpm shaking, at 45°C and 59°C each (for rescreening also at 65°C). The reaction was quenched by adding 100µl Na_2CO_3 per well. Excitation was performed at 365nm, and fluorescence measured at 450nm. The qualitative test for endocellulolytic activity on carboxymethyl cellulose (CMC) was carried out by spotting the expression supernatants onto CMC-Agarose plates (0.5% CMC, 2% agarose) using a 96-pin stamp. After incubation (2h at 50°C, or 65°C) the agar plates were stained with Congo-red, destained with 0.5M NaCl, and photographed.

4.2.9. Data Evaluation

Activities of mutated *Hp*Cel7B were compared to the activity of wild type *Hp*Cel7B (wt*Hp*Cel7B) used as a positive control on the same plate. These values (A_{mut}/A_{wt}) are referred to as index values for the given temperature (45°C, 59°C, or 65°C) (Equation 1). As a final measure of thermostability improvement, while normalizing for possible expression effects due to mutations or varying copy numbers, we used "index ratio", a comparison between index ratios at elevated (59°C or 65°C) and optimal temperature (45°C) (Equation 2). In equation 1 and 2, value A is the measured activity at a certain temperature (T) in RFU (relative fluorescence units). T^{opt} stands for the optimal temperature (45°C), while T^{+} stands for the elevated temperature (59°C, or 65°C).

Equation 1

$$index\ value\ (T) = \frac{A_{clone}(T)}{A_{wt}(T)}$$

Equation 2

$$index\ ratio = \frac{\frac{A_{clone}\,(T^{+})}{A_{wt}\,(T^{+})}}{\frac{A_{clone}\,(T^{opt})}{A_{wt}\,(T^{opt})}}$$

4.2.10. Gene Recovery and Accessing Mutations

Genomic DNA isolation was done based on the protocol given by [190] (for the random mutagenesis clones), or based on the protocol given by [191] (for recombination clones). Isolated genomic DNA was used as a template for PCR amplification of the *Hp*Cel7B gene, using the Phusion polymerase. The PCR conditions were following: 30sec at 98°C, 30 cycles of [10sec at 98°C, 20sec at 60°C, 35sec at 72°C], and 7min at 72°C. Obtained products were purified and sequenced. The sequences were compared to the wild type *Hp*Cel7B.

4.2.11. Temperature Optimum and Half-Lives

Temperature profiles were determined based on activities on MUL substrate at a temperature range from 23°C to 73°C. The MUL assay was performed as described above, except that the incubation time was 1h. For half-live times, residual activity was measured by MUL assay after incubation of expression supernatants at 70°C for 0 to 7min.

4.2.12. Enzyme Purification

All enzyme variants were expressed from single colonies inoculated into 50ml of BMD1 medium (200mM K-phosphate buffer, pH 6, 2% glucose, 13.4g yeast nitrogen base w/o amino acids (Difco)) in 250ml baffled shake flasks at 150rpm, at 28°C for 60h. The supernatants were harvested by centrifugation at 4°C and 4000g for 10min. Buffer exchange and concentrating of the supernatants to end volumes of 1-3ml was achieved by applying 20ml Centricon centrifugation units with a molecular weight cut-off of 10kDa. All enzyme variants were purified by ion exchange chromatography using 1ml DEAE Sepharose columns (HiTrap, GE Healthcare) on an Äkta purifier System with a linear gradient from 0-1M NaCl over 12 column volumes using a 1ml/min flow rate. All elution fractions were tested for glycolytic activity by time-resolved fluorescence measurements (Ex 365nm, Em 450nm) at 25°C using 20µl of each elution fraction and 50µl of a 100µM solution of MUL in 50mM Na-acetate buffer pH 4.8. Activities were determined from linear slopes over a period of 10min. For the final pool of elution fractions only the fractions with the highest fluorescence activities were pooled, the buffer was exchanged to 10mM Na-acetate pH 4.8 and concentrated to 100-300µg/ml (according to UV-absorption) using the Centricon centrifugation units. The final purity of the protein variants was checked by densitometric measurement of the band patterns of SDS-PAGE gels, and the final concentration was measured by Nanodrop 2000.

4.2.13. Thermal Unfolding

All unfolding experiments were performed in 96-well PCR plates using the Applied Biosystems 7500 real-time PCR system. Directly prior to the measurement a 1000-fold dilution of the commercially available Sypro Orange fluorescence dye (Molecular Probes, Life Technologies) was prepared in 10mM Na-acetate buffer, pH 4.8. For the unfolding measurements 5µl of the Sypro Orange dilution were added to 45µL of the protein samples, or buffer-only blanks. In addition to the dye-blanks also blank values for each protein (protein and buffer, without Sypro Orange) were included in the measurement. The plates were thoroughly covered with optically clear adhesive foils, the dye/sample mixtures were mixed for 1min on a Titramax vibration shaker in the dark and after shortly spinning down the samples at 4000xg, and the PCR plate was put into the real-time PCR machine. The increase of the fluorescence signal upon protein unfolding was monitored on-line during a melt-curve analysis with a set heating ramp of 0.5% from 25 to 95°C after initial equilibration at 25°C for 2min. To determine the apparent melting temperatures (T_m) of the protein variants the raw fluorescence data were exported to Origin software, then corrected for the dye and protein blank controls, and finally fitted to a modified Clarke-Fersht equation (Equation 3) [192], with manual setting of the initial values of the pre and post unfolding transition baselines. In the equation, F is the measured fluorescence intensity, dn and kn are the abscissa and slope of pre-unfolding baseline, du and ku are the abscissa and slope of post-unfolding baseline, and T and T_m are the temperature and apparent melting temperature of the protein variant, respectively.

Equation 3

$$F = \frac{dn + kn * T + (du + ku * T) * e^{m*(T-T_m)}}{1 + (du + ku) * e^{m*(T-T_m)}}$$

4.3. Results

4.3.1. Expression in *P. pastoris*

Even though heterologous expression of GH7 family enzymes is believed to be challenging [167], [193] we faced no problems with functional expression of *Hp*Cel7B. Four different LECs were created in order to test the expression and secretion of our target gene under control of two different promoters (inducible AOX1 promoter and constitutive GAP promoter), and two different secretion signals (the native one, and the secretion signal of *Saccharomyces cerevisiae* α-mating factor). To evaluate the

expression/secretion levels, we used the MUL-assay to screen 88 clones of each of the four constructs and create an activity landscape for each of them. By comparing the obtained activity landscapes, the construct containing the *Hp*Cel7B gene with the native secretion signal under control of the GAP promoter stood out as the most active one (data not shown). This construct was therefore chosen as a basis for library creation and further manipulations. Also, a single clone showing average activity within the landscape of the chosen construct was selected as a wild type reference strain.

4.3.2. Random Mutagenesis and Screening

Around 6000 *Pichia pastoris* clones carrying the randomly mutated *Hp*Cel7B genes, generated by error prone PCR with Taq polymerase, were screened employing the MUL activity assay. About 50% of the transformants retained endoglucanase activity. By ranking the obtained index ratios, we selected the 88 most improved clones, which were then rescreened. In the CMC activity assay most of these clones showed preserved activity on this derivative of this polymeric substrate. In a further examination of these 88 clones from the pre-screening we determined the temperatures at half maximal activity (T_{50}). The maximal measured improvement in comparison to the wild type was an increase of 4.09°C (data not shown).

As a final result of this stage, sequencing of the mutated *Hp*Cel7B genes of the rescreening clones revealed an average mutation rate of 2.68 nucleotides per gene, and pointed out several different mutations potentially important for the thermostability. Mutation M63I occurred in four different clones, and two of those clones had the highest index ratios (1.69 and 1.71) in the initial screening of the whole random mutagenesis library. A clone carrying the mutation Q23L had a similar high index ratio value of 1.66. Mutations at position N69 occurred twice among the rescreening clones (N69Y, N69D). Both of these clones also showed improved activity at elevated temperature. A clone carrying mutation G89S had the highest T_{50} value (59.77°C; Δ4.09°C comparing to the wild type).

4.3.3. Structure Guided Design

12 amino acid exchanges were pinpointed as possibly beneficial for the thermostability, based on the created *Hp*Cel7B 3D model and its comparison to the structures of more thermostabile Cel7B endoglucanases. Mutations T29C and G96C, as well as T37C and S109C, were suggested in order to create two enzyme variants carrying an additional disulphide bridge. Disulphide bridges were designed to stabilize highly flexible regions, indicated by high values of predicted B-factors. Another two enzyme variants were created in analogy with more thermostabile homologs *Fo*EG1

and *Hi*EG1, by introducing N-glycosylation sites at position 254 (mutations G254N and P256S), and at position 107 (S107N).

Based on further comparison with *Fo*EG1 and *Hi*EG1 also mutations S53Y, A238Y, N343W, and N373V were suggested as potentially beneficial. These four mutations are all located within surface cavities, and are replacing smaller residues with larger and/or more hydrophobic ones. Increased hydrophobicity of surface cavity residues repels water, thereby possibly contributing to higher thermostability [183]. Overview of all positions targeted for mutation, as well as alignment of *Hp*Cel7B with above-mentioned homologues, are shown in **Figure 10**.

Activity screening using MUL as a substrate revealed that all 8 created *Hp*Cel7B variants retained the enzymatic activity. Variants carrying mutations S107N and N373W showed increased thermostability. Other than these, the clone carrying mutation S53Y gave promising results regarding individual index values at optimal and elevated temperature. The variant carrying mutations T37C and S109C, designed to form a disulphide bridge, showed no change of activity at elevated temperature. Double mutant T29C/G96C was worse. However, keeping in mind a known common destabilisation effect of newly introduced disulphide bridges [194] the neutral effect of mutations T37C/S109C encouraged us to include them in the recombination step to possibly contribute to the thermostability by a synergistic effect.

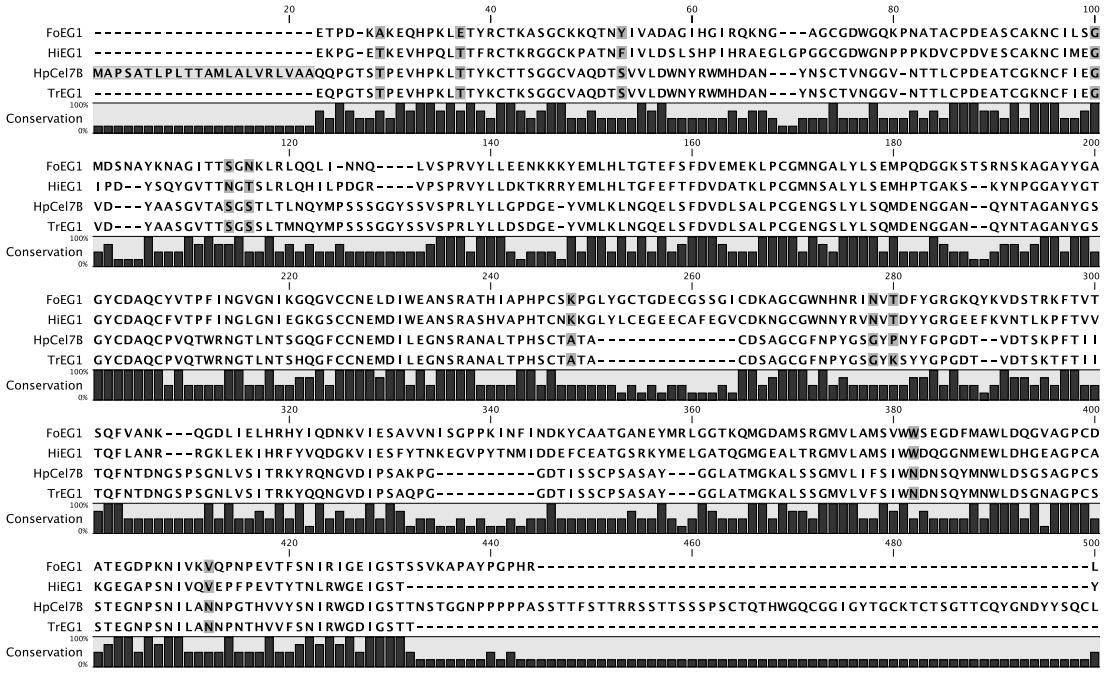

Figure 10 - Sequence alignment of *Hp*Cel7b with its homologues *Fo*EG1, *Hi*EG1, and *Tr*EG1. Signal sequence of *Hp*Cel7B is highlighted in light grey. The mutations targeted for mutagenesis are highlighted in dark grey. Level of conservation is shown as bars below the sequences.

4.3.4. Recombination Library

In order to achieve possible synergistic effects of the thermostability-improving point mutations, we recombined different 9 beneficial mutations by combinatorial gene synthesis using specific codons for the desired mutations: The three best mutations from random mutagenesis (Q23L, N69Y, G89S), 5 coming from the rational design experiments (T37C, S53Y, S107N, S109C, and N373V), and one mutation yielding from both (M63I). By combining 9 different specific triplets the number of all possible variants was 512 (2^9), and statistically most of the clones were expected to contain 4-5 mutations. After screening approximately 1800 clones (3.5-fold library coverage) with the MUL activity assay, about one third of the clones showed improvements. After further selection and rescreening steps, we chose the 12 most improved clones to be examined in more detail.

The *Hp*Cel7B gene was amplified from these 12 clones and sequenced, and the results are shown in **Table 1**. As expected some sequences had an identical set of mutations. Altogether we identified 7 different mutation combinations in these 12 selected clones. All of the finally selected and sequenced clones contained the M63I mutation.

Table 1 - Overview of sequencing results for 12 final *Hp*Cel7B clones resulting from recombination library screening

position	23	37	57	63	69	89	107	109	373	number of mutations
wild type	Q	T	S	M	N	G	S	S	N	
D1				I			N			2
B8				I		S	N			3
B3				I		S	N		V	4
A8,C12	L		Y	I		S	N			5
A10,F12	L		Y	I		S	N		V	6
B12,D2,D3,D6	L		Y	I		S	N	C		6
E5	L	C	Y	I		S		C		6

4.3.5. Mutants Characterisation

We determined loss of activity in response to increasing temperature for 12 selected clones and compared them with the wild type *Hp*Cel7B. A maximal improvement of almost 7°C in terms of T_{50} was observed. The activity decrease of the best two clones over time at 70°C showed improvements of the half-live times ($t_{1/2}$) of up to 1.7 fold (clone B8, **Figure 11** and D1, **Figure 12**) in comparison to the wild type *Hp*Cel7B.

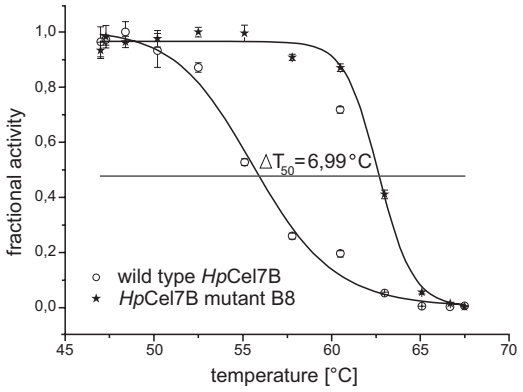

Figure 11 – Temperature profiles of the most improved *Hp*Cel7B variant in terms of T_{50} (namely clone B8, black stars), and a comparison with the wild type (white dots), showing an improvement in temperature at half-maximal activity of almost 7°C. All measurements were done in triplicates.

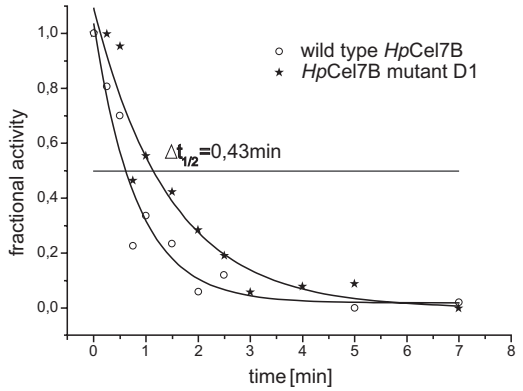

Figure 12 – Activity decay at 70°C in dependence of time for the most improved *Hp*Cel7B variant in terms of half-lives (namely clone D1, black stars), and a comparison with the wild type (white dots), showing an improvement of 0.43min. The results were obtained by single measurements.

In order to determine the denaturation temperature of the finally selected variants, we performed a thermal unfolding assay with purified enzymes. Based on densitometric measurement of the band patterns of SDS-PAGE gels, the final purity of the protein variants was ranging from 25.4% to 64.6% (53.2% on average). The final specific concentrations were derived from this data and measured total protein concentrations, and they were ranging from 80.9µg/ml to 98.0µg/ml (87.5 µg/ml on average) (data not shown). The thermal denaturation was monitored using a real-time PCR machine. All analyzed protein variants showed approximately 5 to >10°C higher melting temperatures in comparison to the wild type enzyme (**Figure 13**). The measured melting temperatures showed strong correlation with estimated T_{50} values, with coefficient of determination (R^2) being 0.93 (data not shown).

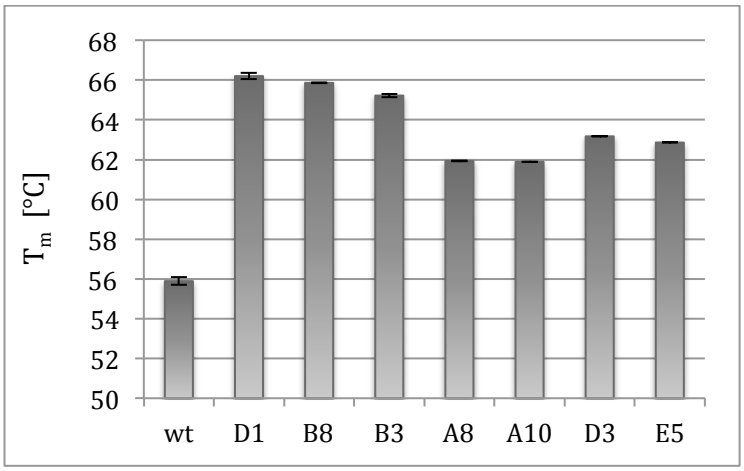

Figure 13 - Apparent thermal melting temperatures T_m, as determined by thermal unfolding of the *Hp*Cel7B variants. All measurements were done in triplicates.

4.4. Discussion

The endoglucanase Cel7B from *Hypocrea pseudokoningii* (*Hp*Cel7B) represented an interesting starting point for improvement of its activity at high temperature. It was necessary, however, to build sufficient knowledge on the enzyme hotspots to identify the amino acid positions mostly contributing to its thermostability and activity. The random mutagenesis round gave a broad overview of mutations that might be playing a major role in *Hp*Cel7B thermostability, and four of those were individually selected for the recombination round (Q23L, M63I, N69Y, G89S). To improve the chances of success in improving *Hp*Cel7B thermostability and identify other beneficial mutations, we analyzed the *Hp*Cel7B 3D structural model and compared it to known structures of other more thermostable GH family 7 endoglucanases. Generally 3 principles were followed.

First, we introduced additional disulphide bonds, to stabilize highly flexible regions: two additional S-S bridges were designed between positions 29 and 96 (T29C, G96C) and between positions 37 and 109 (T37C, S109C). Mutant T37C/S109C showed no improvement of the thermostability, but no deterioration either. As protein destabilization can be a common counter-effect of introducing disulphide bridges [194], this neutral effect encouraged us to include these mutations in the recombination step and allow a possible synergistic effect towards greater thermostability. Second, we introduced additional N-glycosylation sites in the sequence: based on the homolog enzymes *Fo*EG1 and *Hi*EG1, we introduced an N-glycosylation site to *Hp*Cel7B at position 254 (G254N, P256S), and inspired by *Hi*EG1, another one at position 107 (S107N). The latter variant showed improved thermostability, and was included in the recombination step.

Finally, we engineered the entropic energy of the *Hp*Cel7B model: exchanges of smaller to larger and/or more hydrophobic amino acids were suggested to dissipate water from surface cavities. With this aim, and by comparison with *Fo*EG1, we exchanged serine at position 53 to tyrosine (S53Y). Asparagine residues at positions 343 and 373 in *Hp*EG1 were exchanged for tryptophan and valine respectively (N343W, N373V), as these mutations are present in both *Fo*EG1 and *Hi*EG1. Further following the example of *Fo*EG1 and *Hi*EG1, also alanine at position 238 was mutated towards a bulkier amino acid. Instead of lysine, present at analogue position in *Fo*EG1 and *Hi*EG1, we introduced tyrosine (A238Y) in order to increase hydrophobicity and maximize water repulsion from the surface cavity. Out of these, mutant S53Y and mutant N373W gave promising results, and these variants were later included in the recombination step. Also the stabilizing effect of mutation M63I, arising in multiple clones from random mutagenesis can be explained by the same water repulsion effect. Interestingly an isoleucine is also occurring at the corresponding position in both *Fo*EG1 and *Hi*EG1.

After the final semirational recombination step, 12 clones stood out as most thermostabile throughout several rounds of screening, and were characterized in more detail. The temperature profiles of those clones showed no decrease of activity at lower temperatures. This confirmed the idea that improving thermostability of an enzyme does not necessarily sacrifice its activity at low temperatures, even though natural thermostabile enzymes tend to have weak activities at medium temperatures. Of course, this hypothesis is true as long as the stabilising mutations do not have detrimental effects on the active site of the enzyme [195]. By careful analysis of our 3D mutant models, we observed that all stabilizing mutations are indeed on the opposite side of the active site cavity and do not affect its structure (**Figure 14**). Also, this finding confirms the opinion that surface residues have a great influence on protein stability as the surface-located parts are the first to undergo unfolding during thermal denaturation [196], [197].

Considering that the hydrophobic effect is believed to have a central role in protein folding, it can be assumed that the hydrophobic effect is also the major force responsible for protein stability [198]. Also, compactness of a protein is considered to be one of the crucial factors of high stability [199]. This is in accordance with our results, showing that water dissipation from surface cavities, or precisely mutation M63I, most probably has the major effect on the thermostability. This mutation most likely is contributing to the improved thermostability by increasing the hydrophobicity in a surface cavity, thereby stabilizing the loop region. Mutation Q23L, coming from the random mutagenesis approach, is affecting the same cavity, but it seems that no additive effect for high activity at elevated temperature can be generated by combination with M63I.

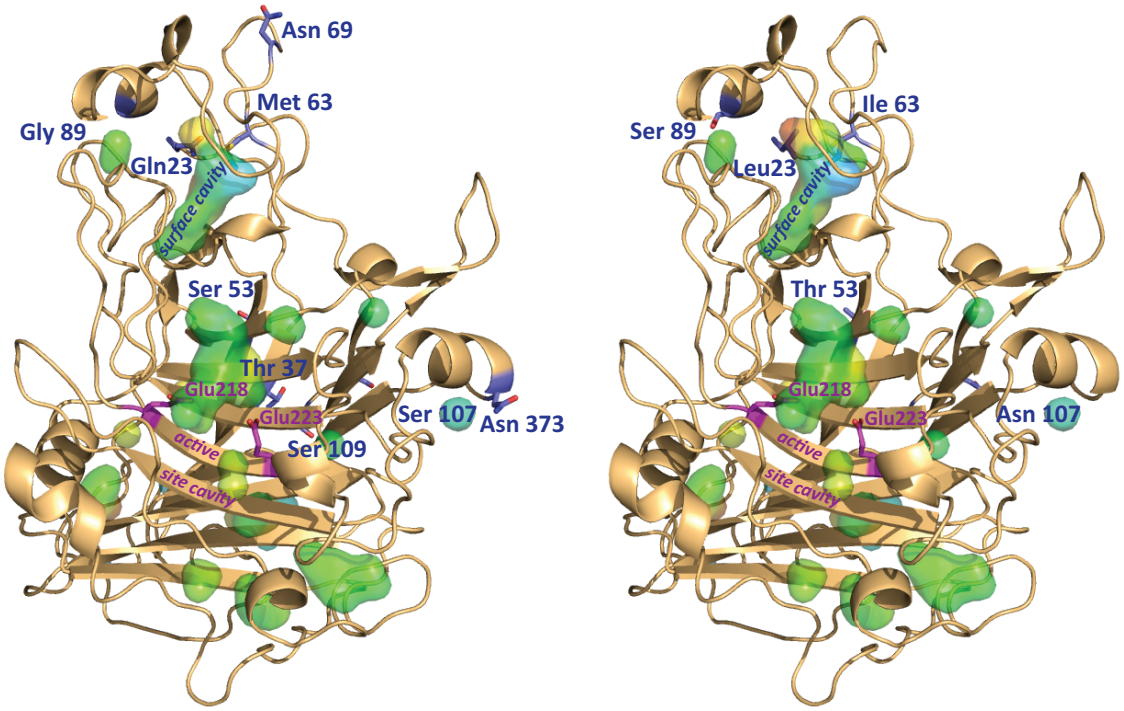

Figure 14 - Comparison of wild type model (left) and one of the best mutant models (A8) (right). Transparent surface representation; the cavities are coloured according to hydrophobicity (red- hydrophobic to blue- hydrophilic). The active site cavity and the surface cavity with the "entropic engineering" effect are indicated accordingly. Proposed active site residues are shown in violet. All residues that were targeted for mutation on the wild type model are shown in stick representation (left blue). The mutated residues of the A8 model are also shown in stick representation (right blue).

4.5. Conclusions

Employing random mutagenesis combined with screening and some structure guided amino acid substitutions the melting temperature of *Hp*Cel7B has been increased by almost 10°C. Mutation M63I seemed to play the central role in stabilizing the enzyme, pointing out the importance of hydrophobic interactions in protein folding and stability. It is generally accepted that thermostability is mostly attained by a small number of single mutations whose effects are additive [200], [201], [202], and those mutations can be derived from structure/sequence knowledge, or a they can be an outcome of a random mutagenesis. In our hands some combinations of beneficial mutations showed additive effects, while others did not further improve activity at high temperature. Random recombination in a smart library made by gene synthesis allowed to identify those useful combinations with limited efforts in screening.

5. Established Protocols

5.1. Preparing and Transformation of Competent *S. pombe* Cells

Purpose and Field of Application

This method describes how to prepare electrocompetent *Schizosaccharomyces pombe* cells suitable for cryopreservation, and how to transform them.

Principle

This method for cryopreservation of intact *S. pombe* cells is based on using the non-permeating cryoprotectant sorbitol, to allow simple and efficient electroporation. It is based on work of *Suga et. al.* [143], [144].

Key Words, Definitions & Abbreviations

Key words: *Schizosaccharomyces pombe, electrocompetent, cryopreservation, transformation*

Abbreviations: *YES media: yeast extract media with amino acid supplements*
YES-Zeo: YES media with 100mg/ml Zeocin
DTT: 1,4-Dithiothreitol

Methodology

Reagents

Name	*Formula*	*MW*	*Purity*	*Supplier*	*Order No.*
Bacto Yeast Extract				*Becton*	*212729*
Bacto Agar				*Becton*	*214030*
Sp supplements				*Formedium*	*PSU0110*
α-D(+)-Glucose Monohydrate	$C_6H_{12}O_6$	*198.17*	≥99.0%	*Carl Roth*	*6780.2*
Zeocin				*InvivoGen*	*ant-zn-5p*
HEPES	$C_8H_{18}N_2O_4S$	*238.31*	≥99.0%	*Carl Roth*	*9105.4*
D-Sorbitol	$C_6H_{14}O_6$	*182.18*		*Carl Roth*	*6213.2*
DTT	$C_4H_{10}O_2S_2$	*154.2*	≥99.0%	*Carl Roth*	*6908.4*

Solutions

YES media: *5g yeast extract, 1.25g/l of S. pombe amino acid supplements, 15g agar (if solid media is needed), 850ml dH_2O; autoclave, then add 150ml 20% glucose, and if needed 100mg Zeocin/l*

20% glucose: *200g glucose/l, autoclaved, stored at 4°C*

DTT buffer: *20mM hepes, 0.6M sorbitol, 25mM DTT; filter sterilized, it can be aliquoted and stored at -20°C*

1M and 2M sorbitol: *sterilized by autoclaving, stored at 4°C*

Materials

Name	*Supplier*	*Order No.*
Rotilabor CryoBox	*Carl Roth*	*P895.2*
Electroporation Cuvettes	*Cell Projects*	*EP-102*

Apparatus

Name	Supplier
Centrifuge 5415 R	*Eppendorf*
Centrifuge 5810 R	*Eppendorf*
Avanti Centrifuge J-20 XP	*Beckman Coulter*

Procedure

Competent Cells Preparation: For preparing competent cells, a *S. pombe* preculture is made by inoculating 50ml YES-Zeo, and cultivated at 30°C over night. This preculture is used to inoculate 500ml of YES-Zeo and create the main culture. The main culture should be grown to cell density of 10^7 cells/ml, and before harvesting it should be placed on ice for 15min. Harvesting is done by centrifugation of the culture for 5min at 1600g. The cell pellet is resuspended in 50ml DTT buffer and incubated at 30°C for 15min. The cells are centrifuged again (5min, 1600g), and washed three times with 50ml ice-cold 1M sorbitol. Finally, the pellet is resuspended with 60ml 2M sorbitol, and 300µl aliquots are placed at -80°C in a CryoBox to freeze gradually, and stored for future transformations.

Transformation: For transformation, aliquots should be thawed at 30°C and then centrifuged for 5min at 1600g. The resulting pellet is washed once with ice-cold 1M sorbitol, and finally resuspended in 40µl 1M sorbitol. The cells are transferred in a pre-chilled cuvette (0.2cm electrode gap) and mixed with 100ng DNA. Electric pulse of 2kV should be applied, and 0.5ml of ice-cold 1M sorbitol added immediately after the electroporation. Transformed cells are kept at 4°C for 90min, and after that 0.5ml YES media is added. After that the cells should be incubated at 30°C for another 90min, with 110rpm shaking. Transformations should be plated on YES-Zeo agar plates, which are then incubated at 28°C for three days.

Safety Precautions

Please follow instructions described in “acib-Mitarbeiterleitfaden Gefahrstoff- und Laborordnung”.

Documentation

All observations, protocol deviations and calculations must be recorded in the lab book (see 537/Alex/3, page 537-389).

5.2. Lignocellulolytic Enzyme Discovery by Expressing Fungal cDNA Library in *S. pombe*

Purpose and Field of Application

This method is used to exploit fungal transcriptomes in order to gain novel sequences coding for lignocellulolytic enzymes.

Principle

The method is based on expressing a eukaryotic cDNA library in *Schizosaccharomyces pombe*, and using a mix of fluorogenic substrates to detect activity in *S. pombe* supernatants.

Key Words, Definitions & Abbreviations

Key words: *cDNA, library, Schizosaccharomyces pombe, fission yeast, activity screening*

Abbreviations:
MU: methylumbelliferyl
MUAr: 4-methylumbelliferyl-α-L-arabinofuranoside
MUC: 4-methylumbelliferyl-β-D-cellobioside
MUGal: 4-methylumbelliferyl-α-D-galactopyranoside
MUGlc: 4-methylumbelliferyl-β-D-glucopyranoside
MUGh: 4-methylumbelliferyl-β-D-glucuronide hydrate
MUL: 4-methylumbelliferyl-β-D-lactopyranoside
MUM: 4-methylumbelliferyl-β-D-mannopyranoside
MUR: 4-methylumbelliferyl-α-L-rhamnopyranoside
MUX: 4-methylumbelliferyl-β-D-xylopyranoside
MUX_2: 4-methylumbelliferyl-β-D-xylobioside
YES media: yeast extract media with amino acid supplements
YES-Zeo: YES media with 100mg/ml Zeocin

Methodology

Reagents

Name	*Formula*	*MW*	*Purity*	*Supplier*	*Order No.*
MUAf	$C_{15}H_{16}O_7$	*308.28*	*≥90%*	*Sigma*	*M9519*
MUC	$C_{22}H_{28}O_{13}$	*500.45*	*≥98%*	*Sigma*	*M6018*
MUGal	$C_{16}H_{18}O_8$	*338.31*	*≥98%*	*Sigma*	*M7633*
MUGlc	$C_{16}H_{18}O_8$	*338.31*		*Sigma*	*M3633*
MUGh	$C_{16}H_{16}O_9 \cdot xH_2O$	*352.29*	*≥98%*	*Sigma*	*M9130*
MUL	$C_{22}H_{28}O_{13}$	*500.45*	*≥98%*	*Sigma*	*M2405*
MUM	$C_{16}H_{18}O_8$	*338.31*	*≥98%*	*Sigma*	*M0905*
MUR	$C_{16}H_{18}O_7$	*322.31*		*Sigma*	*M8412*
MUX	$C_{15}H_{16}O_7$	*308.28*	*≥98%*	*Sigma*	*M7008*

Name	*Formula*	*MW*	*Purity*	*Supplier*	*Order No.*
*EnzCheck® Ultra Xylanase Assay Kit**				*Invitrogen*	*E33650*
Na-carbonate	Na_2CO_3	*105.99*	*≥99%*	*Sigma*	*S7795*
Sp supplements				*Formedium*	*PSU0110*
Ambion RNAqueuos RNA Isolation Kit				*Life Technologies*	*AM1912*
Bacto Peptone				*Becton*	*211820*
Bacto Agar				*Becton*	*214030*
RNase Away				*Sigma*	*83931*
Agarose				*Biozym*	*840004*
Ethidium bromide	$C_{21}H_{20}BrN_3$	*394.29*		*Carl Roth*	*2218.3*
Loading Dye				*Thermo-Scientific*	*R0611*
mRNA-only Eukaryotic mRNA Isolation Kit				*Epicentre*	*MOP51010*
Mint-Universal cDNA Synthesis Kit				*Evrogen*	*SK002*
Herculase II Fusion Polymerase				*Stratagene*	*600675*
Trimmer cDNA Normalization Kit				*Evrogen*	*NK001*
Quick Ligation Kit				*New England Biolabs*	*M2200S*
NEB 10-beta electrocomp. cells				*New England Biolabs*	*C3020K*

** "Component A" = MUX_2*

Solutions

Substrate mix: *MUAf, MUC, MUGal, MUGlc, MUGh, MUL, MUM, MUR, MUX in 10µM concentration, and 5µg/ml of MUX_2, dissolved in 25mM Na-acetate buffer, pH 4.8 (The substrate mix can be frozen in aliquots, and thawed as needed.)*

Quenching solution: *1M Na_2CO_3*

Materials

Name	*Supplier*	*Order No.*
Bio-assay Dish	*Thermo-Scientific*	*240835*

Apparatus

Name	**Supplier**
QPix2 picking robot	*Genetix*
Centrifuge 5415R	*Eppendorf*
Nano Drop 2000c	*Thermo Scientific*
IKA® ULTRA-TURRAX® disperser	*Sigma*

Procedure

RNA Isolation: The fungal strain is grown on agar containing peptone, with an addition of 2% w/v grinded beech wood (to achieve "inductive conditions" for lignocellulose degradation). The beech wood chips (available from pet stores) were dry-milled to a rough powder using IKA® ULTRA-TURRAX® disperser. Beach wood was added to the media (water containing 3g/l peptone and 15g/l agar) before autoclaving. The media should be thoroughly and constantly stirred while pouring the agar-plates, to ensure even distribution of beech wood particles. For RNA isolation 700mg of the mycelium was scraped off the plate and grinded down to homogenous powder using liquid nitrogen, mortar and a pestle. It is important that the mycelium stays frozen at all time during homogenisation. Before RNA isolation is started, the working bench as well as all equipment must be cleaned by RNase Away solution to ensure the environment is decontaminated from RNAses. Wear laboratory gloves at all times during the isolation, to protect the sample from RNases present at the skin. Working under a sterile hood is recommended, but not compulsory. Total RNA isolation was done using Ambion RNAqueous kit (Life Technologies, Paisley, UK). First, still frozen homogenized mycelium powder was roughly divided in three parts, and each third was mixed with 300µl of the Lysis Buffer. Further RNA isolation was done following the manufacturers manual. An equal volume of 64% ethanol is added to the lysate and mixed by vortexing. The lysate-ethanol mix is applied at the filter column placed in a collection tube (both supplied in the kit). Centrifuge at 16000g in an Eppendorf centrifuge 5415R for one minute, and discard flow-through. Apply 700µl of Wash Solution #1 on the filter column, centrifuge, discard flow-through. Repeat the washing step using 500µl Wash Solution #2, then 500µl Wash Solution #3. After discarding the flow-through centrifuge the empty column for another minute. Place the filter column into a fresh collection tube, and apply 40µl RNase-free water preheated to 80°C. Elute by centrifuging for one minute. Repeat the elution step using 30µl preheated RNase free water, and finally using 10µl. The same collection tube should be used for each elution

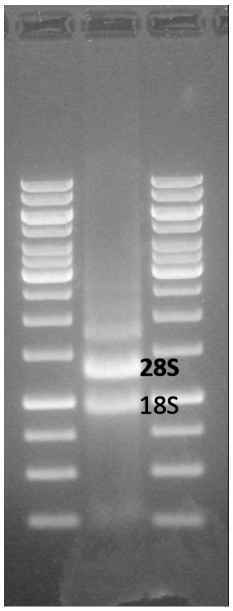

Figure 15 – Example of good-quality total RNA sample (RNA sample in the middle lane, GeneRuler 1kb DNA Ladder in the left and in the right lane). Electrophoresis was done on a 2% agarose gel, and ethidium bromide was used for visualisation. Sharp 28S and 18S bands, and their 2:1 intensity ratio, indicate intact RNA.

step, and around 80µl of eluate is collected per tube. Eluted RNA can be stored at -80°C. The quality of the RNA preparation can be assessed by gel-electrophoresis (2% agarose gel), using ethidium bromide for visualisation. Prior to loading on the gel, the RNA sample is mixed with DNA Loading Dye. An example of good quality RNA is shown in **Figure 15**. Alternatively the quality can be assessed by analysis on an Agilent Bioanalyzer system. RNA concentration can be measured by Nano Drop spectrophotometer.

Library Construction: Companies offering a professional service, such as LGC Genomics or Microsynth, can be hired to do cDNA library construction. LGC Genomics uses the following procedure: For mRNA isolation 10µg of total RNA were used as starting material. Isolation was done using the mRNA-only Eukaryotic mRNA Isolation Kit, with final mRNA purification by LiCl-Precipitation. Synthesis of the cDNA was done by LGC genomics using the Mint-Universal cDNA Synthesis Kit, with 1µg mRNA as starting material. Alteration to the manufacturers instruction manual was using Herculase II Fusion Polymerase for cDNA amplification (for 18 cycles), instead of the Polymerase provided in the Kit. Starting with 600ng of amplified cDNA, the library was normalized using the Trimmer cDNA Normalization Kit. Alteration to the manufacturers instruction manual was using Herculase II Fusion Polymerase for the first amplification of normalized cDNA (for 18 cycles), while the second amplification of normalized cDNA was completely omitted. For library cloning, 2 µg of amplified normalized cDNA was

Figure 16 - Fragment of pGAZ2-SfiI vector (1381-1740), illustrating the modification in respect to the pGAZ2 sequence. Sequence shown in lower case is the ligation stuffer originating from pDNR-lib. Cloning sites of the cDNA library (SfiI A and B) are indicated, as well as the sequence fragment wich gets replaced by a fungal cDNA insert.

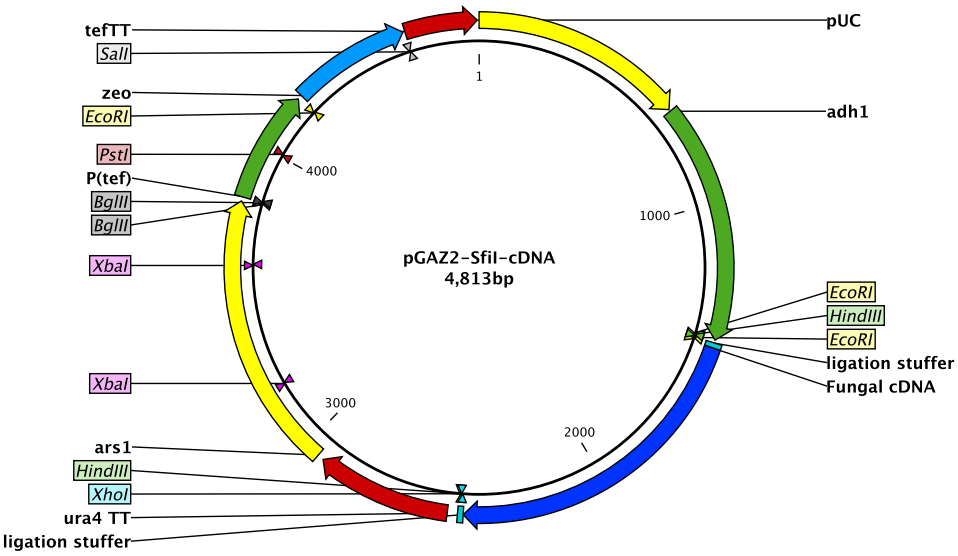

Figure 17 – Vector map of the modified pGAZ2 vector carrying a fungal cDNA insert

digested with 20 Units SfiI for 2h at 48°C. Fragments larger than 800bp were enriched by preparative agarose gel electrophoresis (1.5% agarose). cDNA Fragments were ligated with 100ng pGAZ2-SfiI vector (SfiI cut and dephosporylated) using the Quick Ligation Kit. pGAZ2-SfiI vector was constructed from pGAZ2 and pDNR-lib vector. pGAZ2 was linearized with EcoRV/SpeI. The stuffer fragment of pDNR-lib including both SfiI-sites was released by SmaI/XbaI digestion and moved into pGAZ2 (**Figure 16, Figure 17**) (ligation remark: SmaI and EcoRV are both blunt ends, while XbaI and SpeI have compatible overhangs). Ligation was desalted by precipitation and used to transform NEB 10-beta electrocompetent *E. coli* cells. 4 standard transformations were done in parallel and plated on 8 Large (20cmx20cm) LB-Zeocin plates. More than a million clones were washed from the plates using LB-medium (10 ml per plate). Glycerol was added to a final concentration of 20% and the library was stored at -70°C. The library created in this way was delivered by LGC as an *E. coli* glycerol stock. Plasmids were isolated from this stock, and transformed to *S. pombe* based on the ACIB Protocol "Preparing and Transforming Competent *S. pombe* cells" (see chapter 5.1).

S. pombe Picking and Activity Screening: *S. pombe* colonies were picked to 384 well plates (50µl YES-Zeo/well) using the picking robot (QPix2, Genetix). The 384 well plates were incubated for 2 days at 28°C, in an "aquarium" (glass box saturated with humidity) to prevent evaporation of the media. After this, 30µl 50% glycerol was added to each well, gently shaken for 30 min, and placed at -80°C. These glycerol stocks were then used to inoculate (by stamping) deep-well plates (DWP) filled with 500µl YES-Zeo media/well. The DWPs carrying single *S. pombe* clones from the library were incubated at 30°C, 300rpm, for three days.

Supernatants were collected by 10min centrifugation at 3200rpm. 50µl of the supernatant were mixed with 50µl of the substrate mix, and the sealed black microtiter plates with the reaction mix were incubated for 6h at 37°C with shaking. After the incubation, the reaction was quenched by adding 100µl 1M Na_2CO_3, and fluorescence was measured (λ_{exc}=365nm, λ_{em}=450nm). As a negative control a *S.*

pombe strain carrying the empty pGAZ2 vector was used. Rescreening of active clones with separate MU-substrates was done in the same way like screening, except that 50µl of 100µM (5µg/ml in case of MUX_2) single substrate solutions were used for the assay (all dissolved in 25mM Na-acetate buffer, pH 4.8).
Sequencing: The detected active clones were sequenced using outer multiple cloning site primers for pGAZ2 vector (pGAZ2-MCS-fw: 5'-TCCCATGGTATAAATAGTGGG-3'; pGAZ2-MCS-rv: 5'-GAAAACAAACGCAAACAAGGC-3'). All primers were ordered from Integrated DNA Technologies (IDT) (Coralville IA, USA).

Safety Precautions

Please follow instructions described in "ACIB-Mitarbeiterleitfaden Gefahrstoff- und Laborordnung".

Documentation

All observations, protocol deviations and calculations must be recorded in the lab book (see 120/Alex/4, starting from 120-012).

6. References

[1] **Wyman CE**. 2007. What is (and is not) vital to advancing cellulosic ethanol. *Trends Biotechnol.* **25**: 153–7.

[2] **Solomon S, Plattner G-K, Knutti R, Friedlingstein P**. 2009. Irreversible climate change due to carbon dioxide emissions. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 1704–9.

[3] **Montzka SA, Dlugokencky EJ, Butler JH**. 2011. Non-CO2 greenhouse gases and climate change. *Nature* **476**: 43–50.

[4] **Burgherr P**. 2007. In-depth analysis of accidental oil spills from tankers in the context of global spill trends from all sources. *Journal of hazardous materials* **140**: 245–56.

[5] **Jernelöv A**. 2010. The Threats from Oil Spills: Now, Then, and in the Future. *AMBIO* **39**: 353–66.

[6] **Adsul MG, Singhvi MS, Gaikaiwari SA, Gokhale D V**. 2011. Development of biocatalysts for production of commodity chemicals from lignocellulosic biomass. *Bioresource technology* **102**: 4304–12.

[7] **Wen F, Nair NU, Zhao H**. 2009. Protein engineering in designing tailored enzymes and microorganisms for biofuels production. *Current opinion in biotechnology* **20**: 412–9.

[8] **Alvira P, Tomás-Pejó E, Ballesteros M, Negro MJ**. 2010. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource technology* **101**: 4851–61.

[9] **Schubert C**. 2006. Can biofuels finally take center stage? *Nature biotechnology* **24**: 777–84.

[10] **Banerjee S, Mudliar S, Sen R, Giri B,** et al. 2010. Commercializing lignocellulosic bioethanol: technology bottlenecks and possible remedies. *Biofuels, Bioproducts and Biorefining* **4**: 77–93.

[11] **Alizon F, Shooter SB, Simpson TW**. 2009. Henry Ford and the Model T: lessons for product platforming and mass customization. *Design Studies* **30**: 588–605.

[12] **Rosillo-Calle F, Walter A**. 2006. Global market for bioethanol: historical trends and future prospects. *Energy for Sustainable Development* **10**: 20–32.

[13] **Menon V, Rao M**. 2012. Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept. *Progress in Energy and Combustion Science* **38**: 522–50.

[14] **Fargione J, Hill J, Tilman D, Polasky S,** et al. 2008. Land clearing and the biofuel carbon debt. *Science (New York, N.Y.)* **319**: 1235–8.

[15] **Fargione JE, Cooper TR, Flaspohler DJ, Hill J,** et al. 2009. Bioenergy and Wildlife: Threats and Opportunities for Grassland Conservation. *BioScience* **59**: 767–77.

[16] **Sims REH, Mabee W, Saddler JN, Taylor M**. 2010. An overview of second generation biofuel technologies. *Bioresource technology* **101**: 1570–80.

[17] **Zhu JY, Zhuang XS**. 2012. Conceptual net energy output for biofuel production from lignocellulosic biomass through biorefining. *Progress in Energy and Combustion Science* **38**: 583–98.

[18] **Somerville C, Youngs H, Taylor C, Davis SC,** et al. 2010. Feedstocks for lignocellulosic biofuels. *Science (New York, N.Y.)* **329**: 790–2.

[19] **Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS**. 2002. Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiol. Mol. Biol. Rev.* **66**: 506–77.

[20] **Arantes V, Saddler JN**. 2010. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnology for biofuels* **3**: 4.

[21] **Jäger G, Büchs J**. 2012. Biocatalytic conversion of lignocellulose to platform chemicals. *Biotechnology journal* **7**: 1122–36.

[22] **Huang H-J, Ramaswamy S, Al-Dajani W, Tschirner U,** et al. 2009. Effect of biomass species and plant size on cellulosic ethanol: A comparative process and economic analysis. *Biomass and Bioenergy* **33**: 234–46.

[23] **Spatari S, Bagley DM, MacLean HL**. 2010. Life cycle evaluation of emerging lignocellulosic ethanol conversion technologies. *Bioresource technology* **101**: 654–67.

[24] **Gnansounou E, Dauriat A**. 2010. Techno-economic analysis of lignocellulosic ethanol: A review. *Bioresource technology* **101**: 4980–91.

[25] **Leboreiro J, Hilaly AK**. 2011. Biomass transportation model and optimum plant size for the production of ethanol. *Bioresource technology* **102**: 2712–23.

[26] **Thrän D, Kaltschmitt M**. 2007. Competition--supporting or preventing an increased use of bioenergy? *Biotechnology journal* **2**: 1514–24.

[27] **Yang B, Wyman CE**. 2008. Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuels, Bioproducts and Biorefining* **2**: 26–40.

[28] **Limayem A, Ricke SC**. 2012. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. *Progress in Energy and Combustion Science* **38**: 449–67.

[29] **Sun Y, Cheng J**. 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour. Technol.* **83**: 1–11.

[30] **Alvira P, Negro MJ, Ballesteros M**. 2011. Effect of endoxylanase and α-L-arabinofuranosidase supplementation on the enzymatic hydrolysis of steam exploded wheat straw. *Bioresource technology* **102**: 4552–8.

[31] **Hamelinck CN, Hooijdonk G van, Faaij AP**. 2005. Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. *Biomass and Bioenergy* **28**: 384–410.

[32] **Von Sivers M, Zacchi G**. 1996. Ethanol from lignocellulosics: A review of the economy. *Bioresour. Technol.* **56**: 131–40.

[33] **Klein-Marcuschamer D, Oleskowicz-Popiel P, Simmons BA, Blanch HW**. 2012. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and bioengineering* **109**: 1083–7.

[34] **Percival Zhang Y-H, Himmel ME, Mielenz JR**. 2006. Outlook for cellulase improvement: screening and selection strategies. *Biotechnology advances* **24**: 452–81.

[35] **Maki M, Leung KT, Qin W**. 2009. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *International journal of biological sciences* **5**: 500–16.

[36] **Mellitzer A, Weis R, Glieder A, Flicker K**. 2012. Expression of lignocellulolytic enzymes in Pichia pastoris. *Microbial cell factories* **11**: 61.

[37] **Schülein M**. 2000. Protein engineering of cellulases. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1543**: 239–52.

[38] **Lynd LR, van Zyl WH, McBride JE, Laser M**. 2005. Consolidated bioprocessing of cellulosic biomass: an update. *Current opinion in biotechnology* **16**: 577–83.

[39] **Hahn-Hägerdal B, Galbe M, Gorwa-Grauslund MF, Lidén G,** et al. 2006. Bio-ethanol--the fuel of tomorrow from the residues of today. *Trends in biotechnology* **24**: 549–56.

[40] **Yeoman CJ, Han Y, Dodd D, Schroeder CM,** et al. 2010. Thermostable enzymes as biocatalysts in the biofuel industry. *Advances in applied microbiology* **70**: 1–55.

[41] **Juodeikiene G, Basinskiene L, Vidmantiene D, Makaravicius T,** et al. 2011. The use of β-xylanase for increasing the efficiency of biocatalytic conversion of crop residues to bioethanol. *Catal. Today* **167**: 113–21.

[42] **Kumar R, Wyman CE**Effects of cellulase and xylanase enzymes on the deconstruction of solids from pretreatment of poplar by leading technologies. *Biotechnology progress* **25**: 302–14.

[43] **Jordan DB, Bowman MJ, Braker JD, Dien BS,** et al. 2012. Plant cell walls to ethanol. *The Biochemical journal* **442**: 241–52.

[44] **Margeot A, Hahn-Hagerdal B, Edlund M, Slade R,** et al. 2009. New improvements for lignocellulosic ethanol. *Current opinion in biotechnology* **20**: 372–80.

[45] **Polizeli MLTM, Rizzatti ACS, Monti R, Terenzi HF,** et al. 2005. Xylanases from fungi: properties and industrial applications. *Applied microbiology and biotechnology* **67**: 577–91.

[46] **Žifčáková L, Baldrian P**. 2012. Fungal polysaccharide monooxygenases: new players in the decomposition of cellulose. *Fungal Ecology* **5**: 481–9.

[47] **Cowan D a, Arslanoglu a, Burton SG, Baker GC,** et al. 2004. Metagenomics, gene discovery and the ideal biocatalyst. *Biochemical Society transactions* **32**: 298–302.

[48] **Streit WR, Daniel R, Jaeger K-E**. 2004. Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms. *Current opinion in biotechnology* **15**: 285–90.

[49] **Riesenfeld CS, Schloss PD, Handelsman J**. 2004. Metagenomics: genomic analysis of microbial communities. *Annual review of genetics* **38**: 525–52.

[50] **Simon C, Daniel R**. 2009. Achievements and new knowledge unraveled by metagenomic approaches. *Applied microbiology and biotechnology* **85**: 265–76.

[51] **Tuffin M, Anderson D, Heath C, Cowan DA**. 2009. Metagenomic gene discovery: how far have we moved into novel sequence space? *Biotechnology journal* **4**: 1671–83.

[52] **Venter JC, Remington K, Heidelberg JF, Halpern AL,** et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science (New York, N.Y.)* **304**: 66–74.

[53] **Kalyanaraman C, Imker HJ, Fedorov AA, Fedorov E V,** et al. 2008. Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening. *Structure (London, England : 1993)* **16**: 1668–77.

[54] **Bell PJL, Sunna A, Gibbs MD, Curach NC,** et al. 2002. Prospecting for novel lipase genes using PCR. *Microbiology* **148**: 2283–91.

[55] **Cowan D, Meyer Q, Stafford W, Muyanga S,** et al. 2005. Metagenomic gene discovery: past, present and future. *Trends in biotechnology* **23**: 321–9.

[56] **Meyer QC, Burton SG, Cowan DA**. 2007. Subtractive hybridization magnetic bead capture: a new technique for the recovery of full-length ORFs from the metagenome. *Biotechnology journal* **2**: 36–40.

[57] **LeCleir GR, Buchan A, Maurer J, Moran MA,** et al. 2007. Comparison of chitinolytic enzymes from an alkaline, hypersaline lake and an estuary. *Environmental microbiology* **9**: 197–205.

[58] **Diatchenko L, Lau Y, Campbell AP, Chenchik A,** et al. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 6025.

[59] **Lutz S, Patrick WM**. 2004. Novel methods for directed evolution of enzymes: quality, not quantity. *Current opinion in biotechnology* **15**: 291–7.

[60] **Yuan L, Kurek I, English J, Keenan R**. 2005. Laboratory-directed protein evolution. *Microbiology and molecular biology reviews : MMBR* **69**: 373–92.

[61] **Savile CK, Janey JM, Mundorff EC, Moore JC,** et al. 2010. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science (New York, N.Y.)* **329**: 305–9.

[62] **Chen Z, Zhao H**. 2005. Rapid creation of a novel protein function by in vitro coevolution. *Journal of molecular biology* **348**: 1273–82.

[63] **Bloom JD, Labthavikul ST, Otey CR, Arnold FH**. 2006. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 5869–74.

[64] **Edwards WR, Busse K, Allemann RK, Jones DD**. 2008. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucleic acids research* **36**: e78.

[65] **Heinzelman P, Snow CD, Wu I, Nguyen C,** et al. 2009. A family of thermostable fungal cellulases created by structure-guided recombination. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 5610–5.

[66] **Reetz MT, Carballeira JD, Vogel A**. 2006. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angewandte Chemie (International ed. in English)* **45**: 7745–51.

[67] **Reetz MT, Carballeira JD**. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature protocols* **2**: 891–903.

[68] **Reetz MT, Bocola M, Carballeira JD, Zha D,** et al. 2005. Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angewandte Chemie (International ed. in English)* **44**: 4192–6.

[69] **Jochens H, Aerts D, Bornscheuer UT**. 2010. Thermostabilization of an esterase by alignment-guided focussed directed evolution. *Protein engineering, design & selection : PEDS* **23**: 903–9.

[70] **Jochens H, Bornscheuer UT**. 2010. Natural diversity to guide focused directed evolution. *Chembiochem : a European journal of chemical biology* **11**: 1861–6.

[71] **Höhne M, Schätzle S, Jochens H, Robins K,** et al. 2010. Rational assignment of key motifs for function guides in silico enzyme identification. *Nature chemical biology* **6**: 807–13.

[72] **Wang L, Xie J, Schultz PG**. 2006. Expanding the genetic code. *Annual review of biophysics and biomolecular structure* **35**: 225–49.

[73] **Siegel JB, Zanghellini A, Lovick HM, Kiss G,** et al. 2010. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science (New York, N.Y.)* **329**: 309–13.

[74] **Murphy PM, Bolduc JM, Gallaher JL, Stoddard BL,** et al. 2009. Alteration of enzyme specificity by computational loop remodeling and design. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 9215–20.

[75] **Jiang L, Althoff EA, Clemente FR, Doyle L,** et al. 2008. De novo computational design of retro-aldol enzymes. *Science (New York, N.Y.)* **319**: 1387–91.

[76] **Sterner R, Merkl R, Raushel FM**. 2008. Computational design of enzymes. *Chemistry & biology* **15**: 421–3.

[77] **Yang G, Rich JR, Gilbert M, Wakarchuk WW,** et al. 2010. fluorescence activated cell sorting as a general ultra-high-throughput screening method for directed evolution of glycosyltransferases.

[78] **Becker S, Schmoldt H-U, Adams TM, Wilhelm S,** et al. 2004. Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. *Current opinion in biotechnology* **15**: 323–9.

[79] **Mastrobattista E, Taly V, Chanudet E, Treacy P,** et al. 2005. High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chemistry & biology* **12**: 1291–300.

[80] **Aharoni A, Amitai G, Bernath K, Magdassi S,** et al. 2005. High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chemistry & biology* **12**: 1281–9.

[81] **Collins T, Gerday C, Feller G**. 2005. Xylanases, xylanase families and extremophilic xylanases. *FEMS microbiology reviews* **29**: 3–23.

[82] **Biely P**. 1985. Microbial xylanolytic systems. *Trends Biotechnol.* **3**: 286–90.

[83] **Wong KK, Tan LU, Saddler JN**. 1988. Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications. *Microbiological reviews* **52**: 305–17.

[84] **Dodd D, Cann IKO**. 2009. Enzymatic deconstruction of xylan for biofuel production. *Global change biology. Bioenergy* **1**: 2–17.

[85] **Sharma M, Kumar A**. 2013. Xylanases: An Overview ... *British Biotechnology Journal*

[86] **Saha BC**. 2003. Hemicellulose bioconversion. *Journal of industrial microbiology & biotechnology* **30**: 279–91.

[87] **Shallom D, Shoham Y**. 2003. Microbial hemicellulases. *Current Opinion in Microbiology* **6**: 219–28.

[88] **Ahmed S, Riaz S, Jamil A**. 2009. Molecular cloning of fungal xylanases: an overview. *Applied microbiology and biotechnology* **84**: 19–35.

[89] **Henrissat B, Bairoch A**. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical journal* **293 (Pt 3**: 781–8.

[90] **Törrönen A, Rouvinen J**. 1997. Structural and functional properties of low molecular weight endo-1,4-β -xylanases. *J. Biotechnol.* **57**: 137–49.

[91] **Subramaniyan S, Prema P**. 2002. Biotechnology of microbial xylanases: enzymology, molecular biology, and application. *Critical reviews in biotechnology* **22**: 33–64.

[92] **Beliën T, Van Campenhout S, Robben J, Volckaert G**. 2006. Microbial endoxylanases: effective weapons to breach the plant cell-wall barrier or, rather, triggers of plant defense systems? *Molecular plant-microbe interactions : MPMI* **19**: 1072–81.

[93] **Luo H, Yang J, Li J, Shi P,** et al. 2010. Molecular cloning and characterization of the novel acidic xylanase XYLD from Bispora sp. MEY-1 that is homologous to family 30 glycosyl hydrolases. *Applied microbiology and biotechnology* **86**: 1829–39.

[94] **Flint HJ, Martin J, McPherson CA, Daniel AS,** et al. 1993. A bifunctional enzyme, with separate xylanase and beta(1,3-1,4)-glucanase domains, encoded by the xynD gene of Ruminococcus flavefaciens. *Journal of bacteriology* **175**: 2943–51.

[95] **Chang L, Ding M, Bao L, Chen Y,** et al. 2011. Characterization of a bifunctional xylanase/endoglucanase from yak rumen

microorganisms. *Applied microbiology and biotechnology* **90**: 1933–42.

[96] **Beg QK, Kapoor M, Mahajan L, Hoondal GS**. 2001. Microbial xylanases and their industrial applications: a review. *Appl. Microbiol. Biotechnol.* **56**: 326–38.

[97] **Subramaniyan S, Prema P**. 2000. Cellulase-free xylanases from Bacillus and other microorganisms. *FEMS Microbiol. Lett*. **183**: 1–7.

[98] **Butt MS, Tahir-Nadeem M, Ahmad Z, Sultan MT**. 2008. Xylanases and Their Applications in Baking Industry. *Food Technology and Biotechnology* **46**: 22–31.

[99] **Bhat MK**. 2000. Cellulases and related enzymes in biotechnology. *Biotechnology advances* **18**: 355–83.

[100] **Dhiman SS, Sharma J, Battan B**. 2008. INDUSTRIAL APPLICATIONS AND FUTURE PROSPECTS OF MICROBIAL XYLANASES: A REVIEW. *BioResources* **3**: 1377–402.

[101] **Cleemput G, Bleukx W, van Oort M, Hessing M,** et al. 1995. Evidence for the presence of arabinoxylan hydrolysing enzymes in European wheat flours. *Journal of Cereal Science* **22**: 139–45.

[102] **Yamaura I, Koga T, Matsumoto T, Kato T**. 1997. Purification and some properties of endo-1,4-beta-D-xylanase from a fresh-water mollusc, Pomacea insularus (de Ordigny). *Bioscience, biotechnology, and biochemistry* **61**: 615–20.

[103] **Kulkarni N, Shendye A, Rao M**. 1999. Molecular and biotechnological aspects of xylanases. *FEMS Microbiol. Rev.* **23**: 411–56.

[104] **Collins T, Meuwis M-A, Stals I, Claeyssens M,** et al. 2002. A novel family 8 xylanase, functional and physicochemical characterization. *The Journal of biological chemistry* **277**: 35133–9.

[105] **Wu S, Liu B, Zhang X**. 2006. Characterization of a recombinant thermostable xylanase from deep-sea thermophilic Geobacillus sp. MT-1 in East Pacific. *Applied microbiology and biotechnology* **72**: 1210–6.

[106] **Hurlbert JC, Preston JF**. 2001. Functional characterization of a novel xylanase from a corn strain of Erwinia chrysanthemi. *Journal of bacteriology* **183**: 2093–100.

[107] **Wang G, Luo H, Wang Y, Huang H,** et al. 2011. A novel cold-active xylanase gene from the environmental DNA of goat rumen contents: direct cloning, expression and enzyme characterization. *Bioresource technology* **102**: 3330–6.

[108] **Jiang Z-Q, Deng W, Li L-T, Ding C-H,** et al. 2004. A novel, ultra-large xylanolytic complex (xylanosome) secreted by Streptomyces olivaceoviridis. *Biotechnol. Lett.* **26**: 431–6.

[109] **Větrovský T, Voříšková J, Snajdr J, Gabriel J,** et al. 2011. Ecology of coarse wood decomposition by the saprotrophic fungus Fomes fomentarius. *Biodegradation* **22**: 709–18.

[110] **Mitrovic A, Glieder A, Flicker K**. 2011. New Enzymes Complementing natural diversity with laboratory-based methods. *Chemistry Today* **29**: 12–6.

[111] **Tuohy MG, Puls J, Claeyssens M, Vrsanská M,** et al. 1993. The xylan-degrading enzyme system of Talaromyces emersonii: novel enzymes with activity against aryl beta-D-xylosides and unsubstituted xylans. *The Biochemical journal* **290 (Pt 2**: 515–23.

[112] **Christakopoulos P, Nerinckx W, Kekos D, Macris B,** et al. 1996. Purification and characterization of two low molecular mass alkaline xylanases from Fusarium oxysporum F3. *J. Biotechnol.* **51**: 181–9.

[113] **Xu J, Takakuwa N, Nogawa M, Okada H,** et al. 1998. A third xylanase from Trichoderma reesei PC-3-7. *Appl. Microbiol. Biotechnol.* **49**: 718–24.

[114] **Wang Y, Zhang H, He Y, Luo H,** et al. 2007. Characterization, gene cloning, and expression of a novel xylanase XYNB from Streptomyces olivaceoviridis A1. *Aquaculture* **267**: 328–34.

[115] **Moukouli M, Topakas E, Christakopoulos P**. 2011. Cloning and optimized expression of a GH-11 xylanase from Fusarium oxysporum in Pichia pastoris. *New biotechnology* **28**: 369–74.

[116] **Podkaminer KK, Guss AM, Trajano HL, Hogsett DA,** et al. 2012. Characterization of xylan utilization and discovery of a new endoxylanase in Thermoanaerobacterium saccharolyticum through targeted gene deletions. *Applied and environmental microbiology* **78**: 8441–7.

[117] **Chung H-J, Park S-M, Kim H-R, Yang M-S,** et al. 2002. Cloning the gene encoding acetyl xylan esterase from Aspergillus ficuum and its expression in Pichia pastoris. *Enzyme Microb. Technol.* **31**: 384–91.

[118] **Zhou J, Huang H, Meng K, Shi P,** et al. 2009. Molecular and biochemical characterization of a novel xylanase from the symbiotic Sphingobacterium sp. TN19. *Applied microbiology and biotechnology* **85**: 323–33.

[119] **Kiyohara M, Sakaguchi K, Yamaguchi K, Araki T,** et al. 2005. Molecular cloning and characterization of a novel beta-1,3-xylanase possessing two putative carbohydrate-binding modules from a marine bacterium Vibrio sp. strain AX-4. *The Biochemical journal* **388**: 949–57.

[120] **Brennan Y, Callen WN, Christoffersen L, Dupree P,** et al. 2004. Unusual microbial xylanases from insect guts. *Applied and environmental microbiology* **70**: 3609–17.

[121] **Mo X, Chen C, Pang H, Feng Y,** et al. 2010. Identification and characterization of a novel xylanase derived from a rice straw degrading enrichment culture. *Applied microbiology and biotechnology* **87**: 2137–46.

[122] **Chanjuan L, Hong Y, Shao Z, Lin L,** et al. 2009. Novel alkali-stable, cellulase-free xylanase from deep-sea Kocuria sp. Mn22. *Journal of microbiology and biotechnology* **19**: 873–80.

[123] **Ko C-H, Tsai C-H, Tu J, Lee H-Y,** et al. 2010. Molecular cloning and characterization of a novel thermostable xylanase from Paenibacillus campinasensis BL11. *Process Biochemistry* **45**: 1638–44.

[124] **Schlacher A, Holzmann K, Hayn M, Steiner W,** et al. 1996. Cloning and characterization of the gene for the thermostable xylanase XynA from Thermomyces lanuginosus. *Journal of biotechnology* **49**: 211–8.

[125] **Luo H, Wang Y, Li J, Wang H,** et al. 2009. Cloning, expression and characterization of a novel acidic xylanase, XYL11B, from the acidophilic fungus Bispora sp. MEY-1. *Enzyme Microb. Technol.* **45**: 126–33.

[126] **Dalboge H**. 1997. Expression cloning of fungal enzyme genes; a novel approach for efficient isolation of enzyme genes of industrial relevance. *FEMS Microbiol. Rev.* **21**: 29–42.

[127] **Margolles-Clark E, Tenkanen M, Nakari-Setala T, Penttila M**. 1996. Cloning of genes encoding alpha-L-arabinofuranosidase and beta- xylosidase from Trichoderma reesei by expression in Saccharomyces cerevisiae. *Appl. Envir. Microbiol.* **62**: 3840–6.

[128] **Margolles-Clark E, Tenkanen M, Luonteri E, Penttilä M**. 1996. Three α-Galactosidase Genes of Trichoderma reesei Cloned by Expression in Yeast. *European Journal of Biochemistry* **240**: 104–11.

[129] **Christgau S, L K, T H, L A,** et al. 1996. Pectin methyl esterase from Aspergillus aculeatus: expression cloning in yeast and characterization of the recombinant enzyme.

[130] **Pauly M, Andersen LN, Kauppinen S, Kofod L V.,** et al. 1999. A xyloglucan-specific endo- -1,4-glucanase from Aspergillus aculeatus: expression cloning in yeast, purification and characterization of the recombinant enzyme. *Glycobiology* **9**: 93–100.

[131] **Müller S, Sandal T, Kamp-Hansen P, Dalbøge H**. 1998. Comparison of expression systems in the yeasts Saccharomyces cerevisiae, Hansenula polymorpha, Klyveromyces lactis, Schizosaccharomyces pombe and Yarrowia lipolytica. Cloning of two novel promoters from Yarrowia lipolytica. *Yeast (Chichester, England)* **14**: 1267–83.

[132] **Baneyx F**. 1999. Recombinant protein expression in Escherichia coli. *Curr. Opin. Biotechnol.* **10**: 411–21.

[133] **Yin J, Li G, Ren X, Herrler G**. 2007. Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *Journal of biotechnology* **127**: 335–47.

[134] **Chen R**. 2012. Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnology advances* **30**: 1102–7.

[135] **De Marco A**. 2009. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microbial cell factories* **8**: 26.

[136] **Wood V, Gwilliam R, Rajandream M-A, Lyne M,** et al. 2002. The genome sequence of Schizosaccharomyces pombe. *Nature* **415**: 871–80.

[137] **Idiris A, Tohda H, Kumagai H, Takegawa K**. 2010. Engineering of protein secretion in yeast: strategies and impact on protein production. *Applied microbiology and biotechnology* **86**: 403–17.

[138] **Celik E, Calık P**. 2012. Production of recombinant proteins by yeast cells. *Biotechnology advances* **30**: 1108–18.

[139] **Petrescu-Danila E, Voicu M, Stanescu R, Stoica B,** et al. 2009. Fission yeast Schizosaccharomyces pombe as a producer and secretor of heterologous proteins. *Romanian Biotechnological Letters* **14**: 4201 – 4210.

[140] **Gruber C**. 2008. New Tools for the Expression and Screening of Fungal Enzymes in Schizosaccharomyces Pombe.

[141] **Andryushkova A**. 2009. Establishing fission yeast as a host for eukaryotic genomic libraries.

[142] **Elisashvili V, Kachlishvili E, Tsiklauri N, Metreveli E,** et al. 2008. Lignocellulose-degrading enzyme production by white-rot Basidiomycetes isolated from the forests of Georgia. *World Journal of Microbiology and Biotechnology* **25**: 331–9.

[143] **Suga M, Isobe M, Hatakeyama T**. 2000. Cryopreservation of competent intact yeast cells for efficient electroporation. *Yeast (Chichester, England)* **16**: 889–96.

[144] **Suga M, Hatakeyama T**. 2003. High-efficiency electroporation by freezing intact yeast cells with addition of calcium. *Current genetics* **43**: 206–11.

[145] **Biely P, Vršanská M, Tenkanen M, Kluepfel D**. 1997. Endo-β-1,4-xylanase families: differences in catalytic properties. *J. Biotechnol.* **57**: 151–66.

[146] **Paës G, Berrin J-G, Beaugrand J**. 2012. GH11 xylanases: Structure/function/properties relationships and applications. *Biotechnology advances* **30**: 564–92.

[147] **Sørensen HP**. 2010. Towards universal systems for recombinant gene expression. *Microbial cell factories* **9**: 27.

[148] **Wilson DB**. 2009. Cellulases and biofuels. *Curr. Opin. Biotechnol.* **20**: 295–9.

[149] **Himmel ME, Ding S-Y, Johnson DK, Adney WS,** et al. 2007. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science* **315**: 804–7.

[150] **Moxley G, Zhu Z, Zhang Y-HP**. 2008. Efficient sugar release by the cellulose solvent-based lignocellulose fractionation technology and enzymatic cellulose hydrolysis. *J. Agric. Food Chem.* **56**: 7885–90.

[151] **Lehmann M, Wyss M**. 2001. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr. Opin. Biotechnol.* **12**: 371–5.

[152] **Turner P, Mamo G, Karlsson EN**. 2007. Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microb. Cell. Fact.* **6**: 9.

[153] **Cereghino JL, Cregg JM**. 2000. Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol. Rev.* **24**: 45–66.

[154] **Liu Z, Pscheidt B, Avi M, Gaisberger R,** et al. 2008. Laboratory evolved biocatalysts for stereoselective syntheses of substituted benzaldehyde cyanohydrins. *ChemBioChem* **9**: 58–61.

[155] **Jaenicke R**. 2000. Stability and stabilization of globular proteins in solution. *J. Biotechnol.* **79**: 193–203.

[156] **Jaenicke R, Böhm G**. 1998. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**: 738–48.

[157] **Turner NJ**. 2009. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**: 567–73.

[158] **Liang C, Fioroni M, Rodríguez-Ropero F, Xue Y,** et al. 2011. Directed evolution of a thermophilic endoglucanase (Cel5A) into highly active Cel5A variants with an expanded temperature profile. *J. Biotechnol.* **154**: 46–53.

[159] **Nakazawa H, Okada K, Onodera T, Ogasawara W,** et al. 2009. Directed evolution of endoglucanase III (Cel12A) from Trichoderma reesei. *Appl. Microbiol. Biotechnol.* **83**: 649–57.

[160] **Lutz S**. 2010. Beyond directed evolution--semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **21**: 734–43.

[161] **Anbar M, Lamed R, Bayer EA**. 2010. Thermostability Enhancement of Clostridium thermocellum Cellulosomal Endoglucanase Cel8A by a Single Glycine Substitution. *ChemCatChem* **2**: 997–1003.

[162] **Anbar M, Gul O, Lamed R, Sezerman UO,** et al. 2012. Improved thermostability of Clostridium thermocellum endoglucanase Cel8A by using consensus-guided mutagenesis. *Appl. Environ. Microbiol.* **78**: 3458–64.

[163] **Yi ZL, Pei XQ, Wu ZL**. 2011. Introduction of glycine and proline residues onto protein surface increases the thermostability of endoglucanase CelA from Clostridium thermocellum. *Bioresour. Technol.* **102**: 3636–8.

[164] **Voutilainen SP, Boer H, Alapuranen M, Jänis J,** et al. 2009. Improving the thermostability and activity of Melanocarpus albomyces cellobiohydrolase Cel7B. *Appl. Microbiol. Biotechnol.* **83**: 261–72.

[165] **Serrano L, Day AG, Fersht AR**. 1993. Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J. Mol. Biol.* **233**: 305–12.

[166] **Yennamalli RM, Rader AJ, Wolt JD, Sen TZ**. 2011. Thermostability in endoglucanases is fold-specific. *BMC Struct. Biol.* **11**: 10.

[167] **Voutilainen SP, Murray PG, Tuohy MG, Koivula A**. 2010. Expression of Talaromyces emersonii cellobiohydrolase Cel7A in Saccharomyces cerevisiae and rational mutagenesis to improve its thermostability and activity. *Protein Eng., Des. Sel.* **23**: 69–79.

[168] **Hayashida S, Yoshioka H**. 1980. The role of carbohydrate moiety on thermostability of cellulases from Humicola isolens YH-8. *Agric. Biol. Chem.* **44**: 481–7.

[169] **Olsen O, Thomsen KK**. 1991. Improvement of bacterial -glucanase thermostability by glycosylation. *Microbiology* **137**: 579–85.

[170] **Meldgaard M, Svendsen L**. 1994. Different effects of N-glycosylation on the thermostability of highly homologous bacterial (1,3-1,4)-beta-glucanases secreted from yeast. *Microbiology* **140**: 159–66.

[171] **Wang C, Eufemi M, Turano C, Giartosio A**. 1996. Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry* **35**: 7299–307.

[172] **Penttilä ME, André L, Lehtovaara P, Bailey M,** et al. 1988. Efficient secretion of two fungal cellobiohydrolases by Saccharomyces cerevisiae. *Gene* **63**: 103–12.

[173] **Domínguez A, Fermiñán E, Sánchez M, González FJ,** et al. 1998. Non-conventional yeasts as hosts for heterologous protein production. *International microbiology : the official journal of the Spanish Society for Microbiology* **1**: 131–42.

[174] **Takada G, Kawaguchi T, Sumitani J, Arai M**. 1998. Expression of Aspergillus aculeatus No. F-50 cellobiohydrolase I (cbhI) and beta-glucosidase 1 (bgl1) genes by Saccharomyces cerevisiae. *Bioscience, biotechnology, and biochemistry* **62**: 1615–8.

[175] **Jeoh T, Michener W, Himmel ME, Decker SR,** et al. 2008. Implications of cellobiohydrolase glycosylation for use in biomass conversion. *Biotechnology for biofuels* **1**: 10.

[176] **Ilmén M, den Haan R, Brevnova E, McBride J,** et al. 2011. High level secretion of cellobiohydrolases by Saccharomyces cerevisiae. *Biotechnology for biofuels* **4**: 30.

[177] **Jaenicke R**. 1996. Protein folding and association: in vitro studies for self-organization and targeting in the cell. *Curr. Top. Cell. Regul.* **34**: 209–314.

[178] **Danson MJ, Hough DW**. 1998. Structure, function and stability of enzymes from the Archaea. *Trends Microbiol.* **6**: 307–14.

[179] **Szilágyi A, Závodszky P**. 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**: 493–504.

[180] **Abraham T, Pil Pack S, Je Yoo Y**. 2005. Stabilization of Bacillus subtilis Lipase A by increasing the residual packing. *Biocatal. Biotransform.* **23**: 217–24.

[181] **Joo JC, Pohkrel S, Pack SP, Yoo YJ**. 2010. Thermostabilization of Bacillus circulans xylanase via computational design of a flexible surface cavity. *J. Biotechnol.* **146**: 31–9.

[182] **Martin A, Sieber V, Schmid FX**. 2001. In-vitro selection of highly stabilized protein variants with optimized surface. *J. Mol. Biol.* **309**: 717–26.

[183] **Kotzia GA, Labrou NE**. 2009. Engineering thermal stability of L-asparaginase by in vitro directed evolution. *FEBS J.* **276**: 1750–61.

[184] **Sandgren M, Gualfetti PJ, Shaw A, Gross LS,** et al. 2003. Comparison of family 12 glycoside hydrolases and recruited substitutions important for thermal stability. *Protein Sci.* **12**: 848–60.

[185] **Sandgren M, Ståhlberg J, Mitchinson C**. 2005. Structural and biochemical studies of GH family 12 cellulases: improved thermal stability, and ligand complexes. *Prog. Biophys. Mol. Biol.* **89**: 246–91.

[186] **Krieger E, Koraimann G, Vriend G**. 2002. Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. *Proteins* **47**: 393–402.

[187] **Hendlich M, Rippmann F, Barnickel G**. 1997. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graphics Modell.* **15**: 359–63, 389.

[188] **Steinkellner G, Rader R, Thallinger GG, Kratky C,** et al. 2009. VASCo: computation and visualization of annotated protein surface contacts. *BMC Bioinformatics* **10**: 32.

[189] **Lin-Cereghino J, Wong WW, Xiong S, Giang W,** et al. 2005. Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. *BioTechniques* **38**: 44–8.

[190] **Hoffman CS, Winston F**. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. *Gene* **57**: 267–72.

[191] **Harju S, Fedosyuk H, Peterson KR**. 2004. Rapid isolation of yeast genomic DNA: Bust n' Grab. *BMC Biotechnol.* **4**: 8.

[192] **Clarke J, Fersht AR**. 1993. Engineered disulfide bonds as probes of the folding pathway of barnase: Increasing the stability of proteins against the rate of denaturation. *Biochemistry* **32**: 4322–9.

[193] **Boer H, Teeri TT, Koivula A**. 2000. Characterization of Trichoderma reesei cellobiohydrolase Cel7A secreted from Pichia pastoris using two different promoters. *Biotechnol. Bioeng.* **69**: 486–94.

[194] **Siadat OR, Lougarre A, Lamouroux L, Ladurantie C,** et al. 2006. The effect of engineered disulfide bonds on the stability of Drosophila melanogaster acetylcholinesterase. *BMC Biochem.* **7**: 12.

[195] **Van den Burg B, Eijsink VGH**. 2002. Selection of mutations for increased protein stability. *Curr. Opin. Biotechnol.* **13**: 333–7.

[196] **Eijsink VGH, Bjørk A, Gåseidnes S, Sirevåg R,** et al. 2004. Rational engineering of enzyme stability. *J. Biotechnol.* **113**: 105–20.

[197] **Chakravarty S, Varadarajan R**. 2002. Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. *Biochemistry* **41**: 8152–61.

[198] **Dill KA**. 1990. Dominant forces in protein folding. *Biochemistry* **29**: 7133–55.

[199] **Berezovsky IN, Shakhnovich EI**. 2005. Physics and evolution of thermophilic adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 12742–7.

[200] **Sadeghi M, Naderi-Manesh H, Zarrabi M, Ranjbar B**. 2006. Effective factors in thermostability of thermophilic proteins. *Biophys. Chem.* **119**: 256–70.

[201] **Vieille C, Zeikus GJ**. 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**: 1–43.

[202] **Tracewell CA, Arnold FH**. 2009. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.* **13**: 3–9.

7. Appendix

Table 2 - Strains used in this study

Strain Designation	Host Organism	Host Strain (MTB Strain Coll. No.)	Plasmid/LEC	Media	ACIB Strain Coll. No.
20G16-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	700
26O4-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	701
29A17-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	702
30N11-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	703
31D4-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	704
31F23-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	705
39H3-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	706
39N11-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	707
A10-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	697
A8-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	696
B12-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	698
B3-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	695
B8-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	694
D1-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	693
E5-Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	699
Ff-30H12	*S. pombe*	wild type (3370)	pGAZ2-30H12[1]	YES-Zeo	711
Ff-30H12-*Ec*	*E. coli*	TOP10F'	pGAZ2-30H12[1]	LB-Zeo	745
Ff-51D23	*S. pombe*	wild type (3370)	pGAZ2-51D23[1]	YES-Zeo	712
Ff-51D23-*Ec*	*E. coli*	TOP10F'	pGAZ2-51D23[1]	LB-Zeo	746
Ff-5I9	*S. pombe*	wild type (3370)	pGAZ2-5I9[1]	YES-Zeo	708
Ff-5I9-*Ec*	*E. coli*	TOP10F'	pGAZ2-5I9[1]	LB-Zeo	747
Ff-8N23	*S. pombe*	wild type (3370)	pGAZ2-8N23[1]	YES-Zeo	709
Ff-8N23-*Ec*	*E. coli*	TOP10F'	pGAZ2-8N23[1]	LB-Zeo	748
Ff-9P23	*S. pombe*	wild type (3370)	pGAZ2-9P23[1]	YES-Zeo	710
Ff-9P23-*Ec*	*E. coli*	TOP10F'	pGAZ2-9P23[1]	LB-Zeo	749
wild type Pro52	*P. pastoris*	CBS7435 Mut^{S} (3445)	GAP-Pro52-Zeo	YPD-Zeo	692

[1] **Figure16, Figure 17**