DOI: 10.1002/anie.201309953

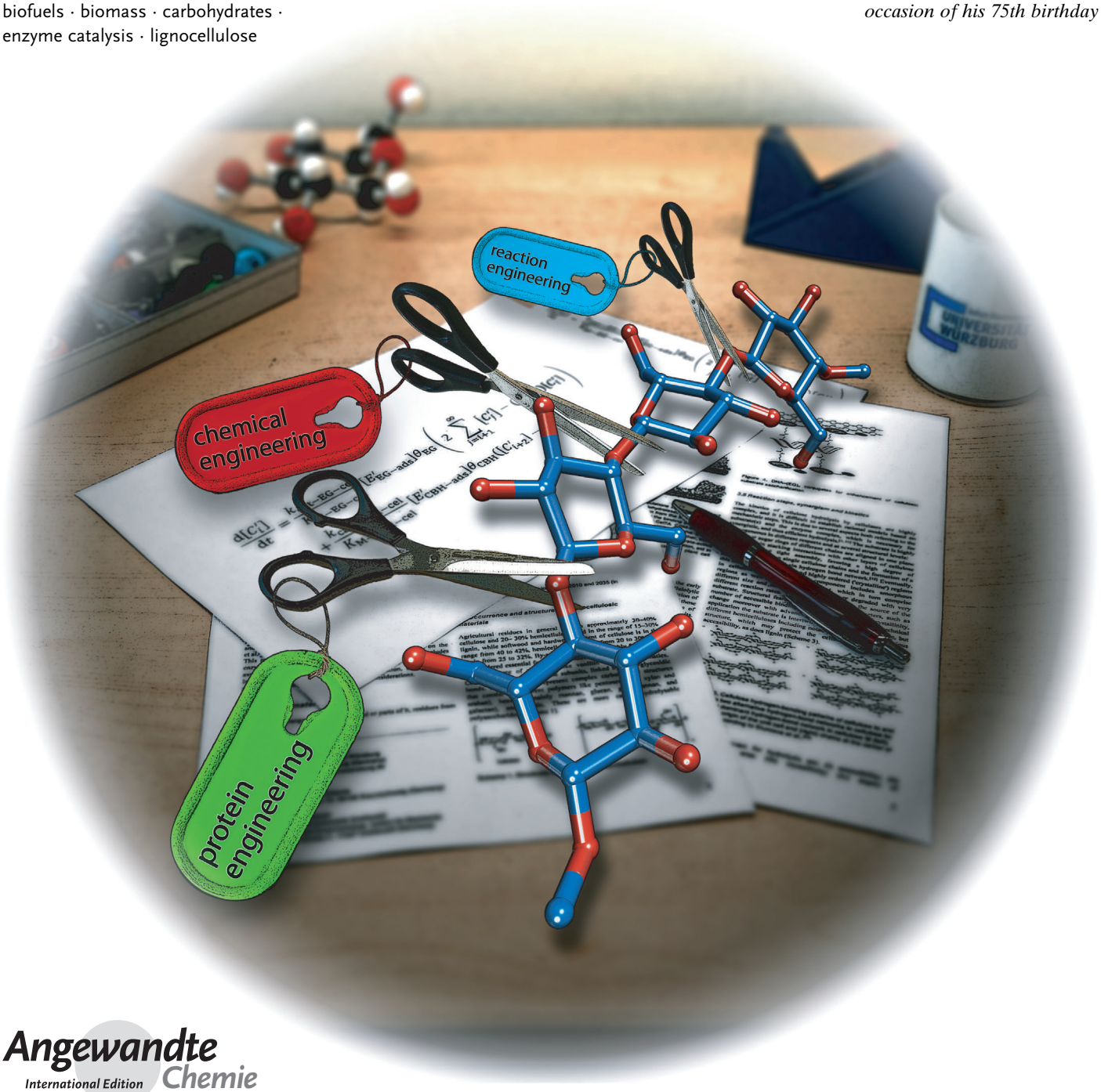
Biofuels

Enzymatic Degradation of (Ligno)cellulose

*Uwe Bornscheuer, Klaus Buchholz, and Jürgen Seibel**

Keywords:
biofuels · biomass · carbohydrates · enzyme catalysis · lignocellulose

Dedicated to Professor Volker Kasche on the occasion of his 75th birthday



Angewandte International Edition Chemie

Glycoside-degrading enzymes play a dominant role in the biochemical conversion of cellulosic biomass into low-price biofuels and high-value-added chemicals. New insight into protein functions and substrate structures, the kinetics of recognition, and degradation events has resulted in a substantial improvement of our understanding of cellulose degradation.

From the Contents

1. Introduction

Biofuels from lignocellulosic material represent a major challenge: They offer huge potential for biofuel manufacture, provide new raw-material sources without competing with food production, and present an opportunity for considerably reduced greenhouse-gas emissions. A life-cycle analysis published recently states that ethanol from cellulose reduces greenhouse gases by 90% as compared to gasoline.[1] However, severe open problems remain to be solved with respect to infrastructure, logistics, and transportation, technical as well as economic problems, competition with other fuels, including biofuels, and notably the combustion of biomass to produce electricity and thermal energy in delocalized installations; also, fracking, as a new competing technology, seems to have again delayed commercial lignocellulose-based projects. Estimates suggest that the cost of producing cellulosic ethanol is about US$ 0.5 per liter, and thus almost twice as high as for starch-based ethanol.[1,2]

Research on enzymatic cellulose degradation has been pursued with varying intensity in the last 30 years.[3] The topic has had several boom cycles since the 1970s, following the first oil crisis. One pioneer was Elwin T. Reese,[4] who investigated the deterioration of cotton tents in the US Army laboratories during the early 1950s. He identified fungi of the genus *Trichoderma* that produced major amounts of cellulases, which were potent in hydrolyzing cellulose to glucose, including one species that was later named *Trichoderma reesei*. Reese was joined in 1956 by Mary Mandels, and the focus of cellulase research changed from the prevention of hydrolysis to its enhancement for the production of glucose. Systematic classification and characterization, as well as early research on the reactions and synergism of cellulases, was pioneered by the research group of Henrissat.[5] Among early reviews and publications concerning kinetics and processing are those by Ladisch and co-workers,[6] Ghose,[7] and Buchholz and co-workers.[8] This review provides an update on the current status of the enzymatic degradation of (ligno-)cellulosic material. It examines the availability and structure of the starting materials, the enzymes required, their mode of action and discovery, protein engineering, and bioprocess considerations.

2. Biomass Materials

A rather broad range of available substrates have been investigated as starting materials, including wood or parts of it, residues from agricultural raw-materials processing in the food industry, crop residues, such as corn stover, corncobs, and sugarcane bagasse, and energy crops, notably switchgrass and hybrid poplar. Biofuels are currently used to generate 10% of the energy produced worldwide, or 53 exajoules.[9] The global bioenergy potential from dedicated biomass plantations in the 21st century has been estimated with regard to a range of sustainability requirements to safeguard food production, biodiversity, and terrestrial carbon storage. Biofuels hold the potential to provide one quarter of the global future energy demand (153 of 623 exajoules in 2035; Figure 1).[10]

2.1. Occurrence and Structures of Lignocellulosic Materials

Agricultural residues in general contain approximately 30–40% cellulose, 20–30% hemicelluloses, and in the range of 15–30% lignin, whereas the softwood and hardwood cellulose content is in the range of 40–42%, roughly 20–30% hemicelluloses and 25–32% lignin. By-product utilization, notably that of lignin, is considered essential for a viable industrial process. Lignin can be a suitable source of chemicals such as vanillin and other aromatic compounds. Cellulose occurs as D-glucose subunits linked by β-1,4-glycosidic bonds. Hemicelluloses occur as complex carbohy-

[*] Prof. Dr. J. Seibel
Institut für Organische Chemie
Julius-Maximilians-Universität Würzburg
Am Hubland, 97074 Würzburg (Germany)
E-mail: seibel@chemie.uni-wuerzburg.de
Prof. Dr. K. Buchholz
Institut für Technische Chemie
Hans-Sommer-Strasse 10, 38106 Braunschweig (Germany)
Prof. Dr. U. Bornscheuer
Ernst-Moritz-Arndt-Universität Greifswald
Biotechnologie und Enzymkatalyse, Institut für Biochemie
Felix-Hausdorff-Strasse 4, 17487 Greifswald (Germany)

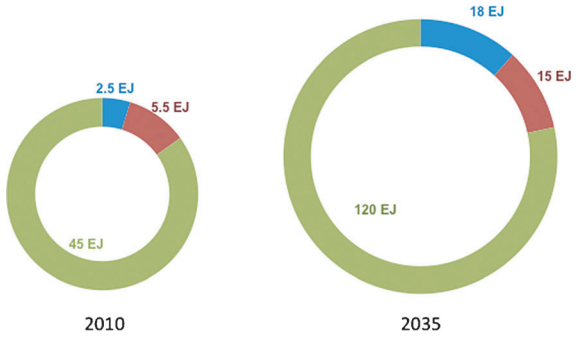
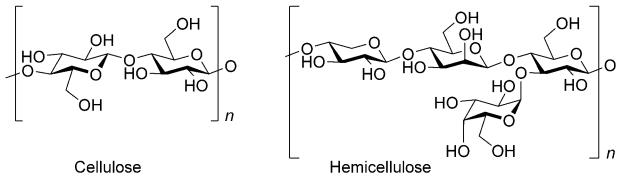


Figure 1. Uses of energy from biomass in 2010 (total 53 EJ) and 2035 (total 153 EJ (exajoules); 10^{18} J). Green: heat; blue: transport; red: electricity.[10]

drate structures composed of different polysaccharides, such as pentoses (e.g., xylan and araban), hexoses (mainly mannan, glucan, glucomannan, and galactan), and pectin. These polysaccharides are readily hydrolyzed (Scheme 1).



Scheme 1. Structures of cellulose and hemicellulose.

In contrast, lignin is a random three-dimensional phenylpropanoid (C_9) polyphenol mainly linked by arylglycerol ether bonds between the monomeric phenolic *p*-coumaryl (**H**), coniferyl (**G**), and sinapyl alcohol (**S**) units (Scheme 2).[11] Gymnosperm lignin consists almost entirely of **G** (G-lignin), dicotyledonous angiosperm lignin is a mixture of **G** and **S** (GS-lignin), and monocotyledonous lignin is a mixture of all three units (GSH-lignin). The polymer is almost completely resistant against microbial (and enzymatic) attack.[12]

2.2. Pretreatment

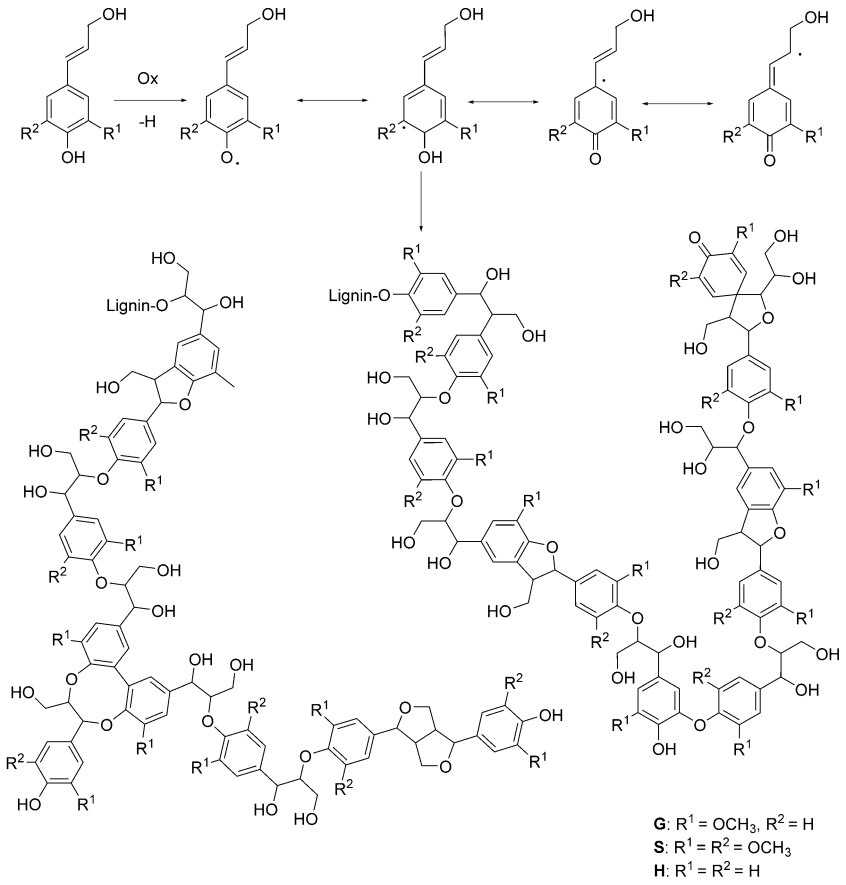
A major issue concerning the access of the enzyme to the sugar components of (ligno-)cellulosic materials the resistance of plant cell walls to deconstruction. This key obstacle for the cost-efficient production of fuels and chemicals needs to be addressed first to make cellulosic solids digestible with yields of enzymatic hydrolysis higher than 90% within reasonable time periods (up to 1 day) so that a high sugar concentration (above 10%, for glucose, additional arabinose, and xylose) results for the subsequent fermentation step. Without pretreatment, enzymatic conversion of the substrate, the reaction rate, and product yields would be very low and therefore uneconomical. The key factors are the dissolution of hemicelluloses and the alteration of the lignin structure to provide improved accessibility of cellulose to hydrolytic enzymes through a deprotected, disrupted, and hydrated structure with increased pore volume. The substrate characteristics that impact the rate of enzymatic hydrolysis include the degree of cellulose crystallinity: Regions of high crystallinity are separated by amorphous regions, surrounded by hemicelluloses, and protected by lignin. The formation of fermentation inhibitors (formic, acetic, and levulinic acids, furan derivatives, and phenolic compounds) must be minimized. Minimum heat and power requirements and low costs are essential. Herein only a short introduction can be given; recent reviews on pretreatment include all relevant aspects and should be studied for details. These aspects are: substrate size reduction and accessibility, the role of component removal, in particular the relative importance of hemicellulose versus lignin removal, component inhibition, reaction conditions, and substrate-specific optimization.[12,13]

The elementary fibrils of cellulose and their aggregations are determined by nature in native fibers, such as cotton or wood-pulp fibers, and are laid down in various cell-wall layers in a typical manner.[14] Three different types of polymers, cellulose, hemicelluloses, and lignin, are intimately associated with one another and inhibit enzymatic degradation.

Leu and Zhu[13c] concluded that hemicellulose removal is more important than lignin removal for creating cellulase-accessible pores. Partial delignification is needed for the satisfactory saccharification of lignocelluloses with a high lignin content, such as those in softwood species. Treatment

Uwe T. Bornscheuer completed his PhD in chemistry at the University of Hannover in 1993. After postdoctoral studies at the University of Nagoya (Japan), he completed his habilitation in 1998 at the Institute of Technical Biochemistry of the University of Stuttgart. He has been Professor at the Institute of Biochemistry at the University of Greifswald since 1999. He is Editor-in-Chief of Eur. J. Lipid Sci. Technol. and a Chairman of ChemCatChem. In 2008, he received the BioCat2008 Award and in 2012 the Chevreul Medal of the French Society of Lipid Research.

Klaus Buchholz studied chemistry at the Universities of Saarbrücken and Heidelberg and received his PhD from the Technical University of Munich in 1969. He subsequently established a research group on enzyme technology at Dechema e.V. in Frankfurt/Main. He completed his habilitation in 1981 at the Technical University of Braunschweig and joined its Institute for Agricultural Technology and the Sugar Industry in 1982. He was head of the Institute from 1988 as well as Professor for the Technology of Carbohydrates at the Institute for Technical Chemistry in Braunschweig from 1991.



Scheme 2. Lignin synthesis and structure.[11] The polymerization step is catalyzed by peroxidase and laccase enzymes by radical coupling reactions between delocalized phenolic monolignol radicals in a combinatorial fashion.

that modifies the surface properties of lignin can be more efficient for reducing or eliminating cellulase inhibition by lignin, thus leading to improved lignocellulose saccharification.

Steam explosion is the most widely employed physicochemical pretreatment for lignocellulosic biomass. The biomass is subjected to pressurized steam at a high temperature (up to 240 °C) for several minutes and then suddenly depressurized.[12] This pretreatment combines mechanical forces due to explosive decompression and chemical effects due to hydrolysis by the protons of water and the acetyl groups present in hemicelluloses (formation of acetic acid). Lignin is redistributed and to some extent removed from the material. The removal of hemicelluloses exposes the cellulose surface and increases enzyme accessibility to the cellulose. The pretreatment leads to good yields in enzymatic hydrolysis with high sugar recovery and avoids the formation of inhibitors. It has been proven efficient for a wide range of raw materials, such as poplar hardwoods and agricultural residues (e.g. corn stover and wheat straw). It is suitable for industrial-scale development, requires lower capital investment, and offers more potential for energy efficiency than other technologies, such as pretreatment with ionic liquids.[12]

Pretreatment with liquid hot water involves the application of water in the liquid state under pressure at elevated temperatures (2–25 bar, 160–240 °C, reaction times ranging from a few minutes up to one hour) to cause alterations in the structure of the lignocelluloses, as induced by the change in pH value of liquid water under these conditions.[12] The objective is to solubilize mainly the hemicellulose, to make the cellulose more accessible, and to avoid the formation of inhibitors (e.g., by control of the pH value). Two fractions are obtained, a cellulose-rich solid, and a liquid fraction rich in hemicellulose oligomers. Examples of raw materials that have been investigated are corn stover, sugarcane bagasse, and wheat straw. In a detailed study, the hydrolysis yield of glucose increased from 3–9 % for the use of natural nonpretreated wheat straw to nearly 100 % under optimum conditions of 200 °C and 20 min with commercial enzyme preparations, including Celluclast (derived from *T. reesei*) and Novozyme 188 (both products from Novozymes, Denmark).[13b] This method is attractive owing to its cost-saving potential and reactor construction (low-corrosion potential).[12,13h,i]

Organosolv-like processing has been long discussed for the delignification of the wood ultrastructure.[15] Organic solvents (ethanol or short-chain alcohols) and water are used to dissolve and thus separate part of the lignin and hemicelluloses. The dissolved lignin and hemicelluloses can then be recovered by changing the concentration of the organic solvent. The high cost of the solvent is regarded as prohibitive in organosolv processes.[16] As an alternative, a biphasic process for biomass pretreatment has been suggested. In this Organocat process, an aqueous solution of oxalic acid is used as a catalytic agent and 2-methyltetrahydrofuran (MTHF) is used as an organic solvent.[17] MTHF solubilizes lignin, whereas the partially hydrolyzed hemicelluloses are dissolved in the aqueous phase. Cellulose remains as a solid. This

Jürgen Seibel studied chemistry at the University of Göttingen and completed his PhD with Prof. Lutz F. Tietze. In 2000, he moved to the University of Oxford for postdoctoral research with Prof. Chris Schofield. In 2002, he began his independent research career at the University of Braunschweig. He moved to the Helmholtz Centre for Infection Research in 2007 and since 2009 has been professor at the University of Würzburg. He was awarded the Jochen Block Prize by Dechema (Society for Chemical Engineering and Biotechnology) in 2008 and the DuPont Young Professor Award in 2012.

process has been analyzed recently for economic viability.[16] The proposed energy consumption is 5 MJ per kg of dry biomass and is mostly due to solvent recycling. The attractive features of the Organocat process are the efficiency of lignin separation and the synthesis of aromatic compounds, propylene, and butadiene.

3. Cellulose-Degrading Enzymes

3.1. CAZyme Classification

Glycoside hydrolases are key enzymes of carbohydrate metabolism that are found in the three major kingdoms (archaebacteria, eubacteria, and eukaryotes). The IUB enzyme nomenclature is based on the type of reaction that enzymes catalyze and on their substrate specificity. For glycoside hydrolases (EC 3.2.1.x), the first three digits are indicative of enzymes that hydrolyze *O*-glycosyl linkages, whereas the last number indicates the substrate and sometimes reflects the molecular mechanism. In 1991, Henrissat classified glycosyl hydrolases on the basis of amino acid sequence similarities and suggested a revision of the enzyme nomenclature.[18] More data was collected, and the database CAZy—"Carbohydrate-Active Enzymes" (http://www.cazy.org/)[19]—was born. The CAZy database describes families of structurally related catalytic and carbohydrate-binding modules (or functional domains) of enzymes that degrade, modify, or create glycosidic bonds. The CAZy database contains 133 families of glycosyl hydrolases[18,20] (glycosidases and transglycosidases), 94 families of glycosyltransferases,[21] 22 families of polysaccharide lyases,[22] and 16 families of carbohydrate esterases. More recently, redox enzymes that act in conjunction with CAZymes were also added as "auxiliary activities".[23] These families are given a number to identify them; for example, Glycosyl hydrolase family 1 contains enzymes that possess a TIM barrel fold. These families are clustered into 14 different clans that share structural similarity. Currently, CAZyme information from 2931 bacterial, 167 archaea, 76 eukaryota, and 255 virus genomes are covered in the database.

CAZy data are accessible either by browsing sequence-based families or by browsing the content of genomes in carbohydrate-active enzymes. New genomes are added regularly shortly after they appear in the daily releases of GenBank. The mission statement of the CAZy database is to cover and classify all carbohydrate-active enzymes across organisms and subfields of the glycosciences.

A typical cellulase system comprises three main activities: chain-end-cleaving cellobiohydrolases (CBHs), internally chain cleaving endoglucanases (EGs), and ß-glucosidases (BGLs), which hydrolyze soluble short-chain glucooligosaccharides to glucose. EGs mainly act on amorphous parts of cellulose. Different CBHs may act on either amorphous or crystalline parts of cellulose. EGs create new chain ends, whereas CBHs proceed with further hydrolysis to produce mainly cellobiose (two glucose units), which is hydrolyzed by BGLs to yield glucose. The different enzyme systems and their diverse interactions with carbohydrate-binding modules (CBMs) may be much more complex, as discussed in Section 3.5.

3.2. Occurrence of CBM

In general, glycoside hydrolases are inefficient enzymes for the degradation of insoluble polysaccharides, as it is difficult for the protein to access the substrate at the right position for catalysis. To overcome these limitations, glycoside hydrolases are often organized in a modular manner. Thus, these systems include the hydrolases as catalytic modules, which are connected to one or more carbohydrate-binding modules (CBMs). These CBMs are the noncatalytic part of many polysaccharide-targeting enzymes, such as cellulases and hemicellulases, and bind to the cell-wall polymers. But how do they function exactly? Recent findings indicate that CBMs enhance enzyme penetration into the cellulose substrates. Gilbert et al. raised important questions in an excellent article: What are the carbohydrate motifs recognized by CBMs, and what are the structural features that drive recognition of these motifs?[24]

The CAZy database groups CBMs in 67 families[25] according to their similar protein sequences, and 54 of those families are so far structurally represented. The CBM modules are further divided into Types A, B, and C. This classification has been introduced to define CBMs in terms of their binding specificity: Type A CBMs bind the surface of crystalline polysaccharides, Type B CBMs recognize internal glycan chains (endo type), and Type C CBMs bind the termini of glycans (exo type).

3.3. Interplay between CBM and Glycoside Hydrolases and Linkers

There are different mechanisms that contribute to the binding between carbohydrates and proteins, represented by CBMs, enzymes, and linkers. One typical mechanism between Type A CBM63 from the β-1,4-(exo)glucanase Cex of *Cellulomonas fimi* and insoluble bacterial microcrystalline cellulose is the entropically driven expulsion of water molecules from hydrophobic surfaces of the cellulose and the protein.[26]

In other mechanisms, we observe a liaison of CBM and the hydrolase for carbohydrate recognition. Xyn10B from *Clostridium thermocellum* is a typical modular enzyme containing an N-terminal family 22 CBM (CBM22-1), a family 10 glycoside hydrolase catalytic domain (GH10), a second CBM22 module (CBM22-2), a dockerin sequence, and a C-terminal family 1 carbohydrate esterase (CE1) catalytic domain. Residues from helix H4 of the GH10 module provide the major contacts by fitting into the minor groove of the CBM22-1 module.[27] The orientation of CBM22-1 is such that it enables the substrate to be loosely bound and subsequently delivered to the active site in a processive manner (Figure 2).[27] Many of these enzymes that degrade lignocellulose are multimodular with CBMs and catalytic domains connected by flexible, glycosylated linkers. These linkers have long been thought to simply serve as

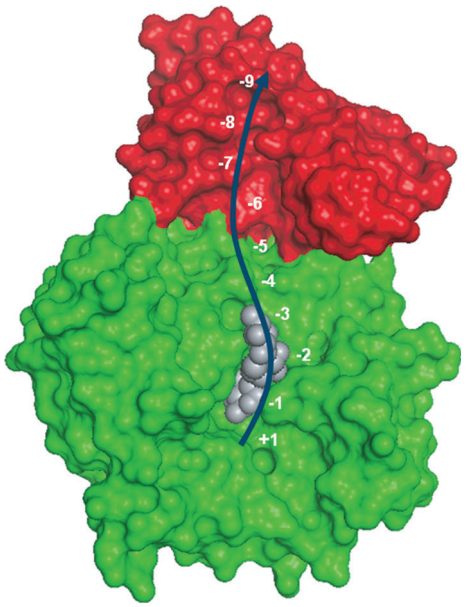


Figure 2. Structure of the putative xylan-binding tract. The CtCBM22-1–GH10 monomer is shown as a solid surface to emphasize that the CBM22-1 (red) and GH10 (green) modules are intimately associated in such a manner that the putative ligand-binding cleft on the CBM22-1 module is aligned with the GH10 catalytic cleft to form a long continuous tract (arrow) that could potentially accommodate xylan chains with at least 10 xylose residues. In the crystal structure (PDB code: 2w5f), the xylotriose moiety is observed in subsites −1 to −3.[27]

a tether between structured domains. Very recently, it was shown that the linker enhances binding over that of the CBM alone by an order of magnitude; it exhibits dynamic binding to the substrate and is thus probably an important contributor to enzyme activity (Figure 3).[28] MD simulations before and after binding of the linker also suggest that the bound linker

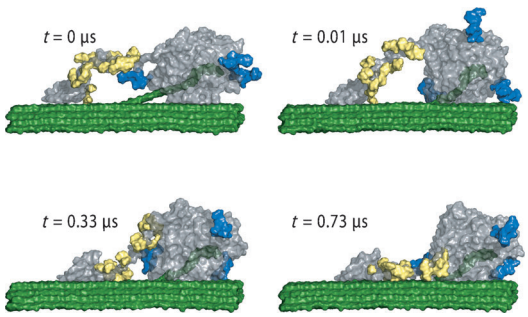


Figure 3. Molecular dynamics (MD) simulations of the *T. reesei* family 6 and family 7 cellobiohydrolases (TrCel6A and TrCel7A) bound to cellulose. The prediction from the MD simulations was examined experimentally by measuring the binding affinity of the Cel7A CBM and the natively glycosylated Cel7A CBM linker. On crystalline cellulose, the glycosylated linker enhances the binding affinity with respect to that observed with the CBM alone by an order of magnitude. Together, these results suggest that glycosylated linkers in carbohydrate-active enzymes, which are intrinsically disordered proteins in solution, aid in dynamic binding during the enzymatic deconstruction of plant cell walls.[28]

may affect enzyme function owing to significant dampening of the enzyme fluctuations.

3.4. Multienzyme Complexes

Cellulosomes are multienzyme complexes discovered in the early 1980s.[29] They were designed by cellulolytic microorganisms as nanomachines for the efficient deconstruction of cellulose and hemicellulose.[30] The "nanomachines" are composed of two complementary structural modules located on two separate types of interacting subunits. The cohesin module on the scaffoldin binds a dockerin module on each enzymatic subunit.[31] On the basis of this principle, various cellulolytic subunits, such as endoglucanases, cellobiohydrolases, xylanases, and other degradative enzymes, work synergistically to degrade heterogeneous, insoluble cellulose substrates. The attachment of the cellulosome to its substrate is mediated by a scaffoldin-borne cellulose-binding module (CBM). The fabrication of artificial cellulosomes to improve the efficiency of enzymatic saccharification has been reported.[32] In a chemical approach, a multicellulase conjugate was assembled on a double-stranded DNA scaffold. The resulting DNA–$(endoglucanase)_n$ conjugate (Figure 4) exhibited unique hydrolytic activity towards crystalline cellulose (Avicel). Its activity depended on the cellulase/DNA ratio of the DNA-based artificial cellulosome.[33]

3.5. Reaction Steps, Synergism, and Kinetics

The kinetics of cellulose hydrolysis by cellulases are highly complex, and it is difficult to establish rational models based on mechanistic steps. This difficulty is due to several aspects with respect to both the substrate(s) complex and the enzyme complex, which is necessary for synergistic, and thus efficient hydrolysis.[3,34] The substrate is highly heterogeneous in different respects and at several structural levels:

The planar sugar rings of different chains are aligned in one plane and engage in stacking interactions with other layers of cellulose chains, thus allowing tight assembly and favoring a high degree of organization between single cellulose chains and the formation of a stable intra- and interchain-hydrogen-bond network.[35] Eventually, these interactions lead to tightly packed and highly ordered ("crystalline") regions in cellulose.[36] The cellulose component includes amorphous regions as well as crystalline regions, which in turn have different sizes and accessibility and thus are degraded with very different reaction rates. These factors vary with the source of the substrate. Moreover, structural features and the relevant parameters, such as the number of accessible binding sites, chain length, and crystallinity, change with advancing degradation.[3,34a] In technical applications, the substrate is intertwined with not only cellulose, but different hemicelluloses, including notably arabinoxylans with various structures, which may protect the cellulose parts and limit accessibility, as does lignin (Scheme 3).

Thus, factors relevant for hydrolysis are: 1) accessibility, 2) available surface area, 3) crystallinity, 4) degree of poly-

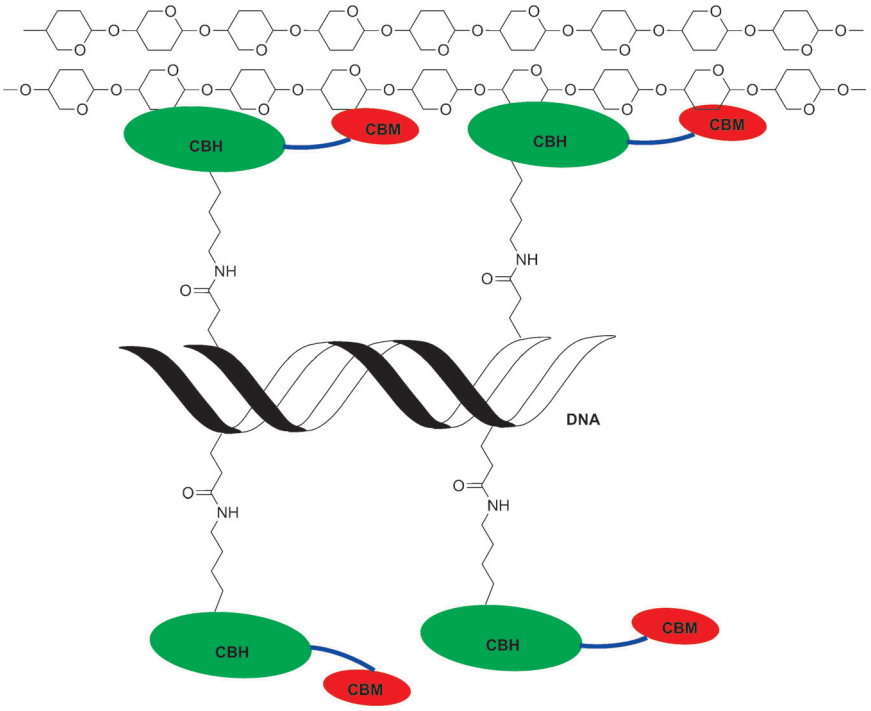


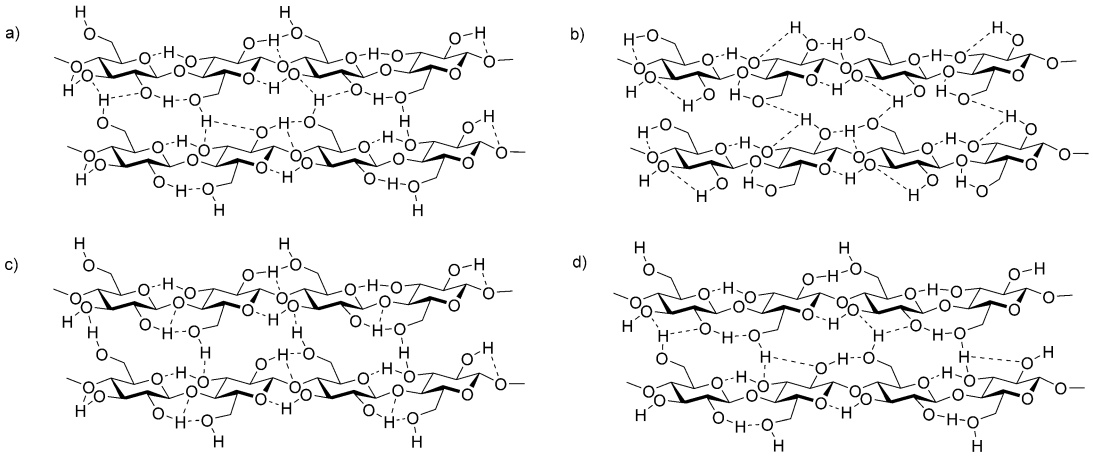
Figure 4. DNA–$(EG)_n$ conjugates for enhancement of the saccharification of cellulosic substrates.

merization, 5) lignin and hemicellulose content, 6) changes in features during degradation, and 7) pretreatment.[37] Adsorption and desorption, as well as unproductive binding of the different enzyme activities on different, heterogeneous substrate sites has to be taken into account.[34b,38] Furthermore, diffusion in the heterogeneous system may play a role.[39] Steps involved in enzyme hydrolysis are: adsorption at the (heterogeneous) substrate surface; lateral diffusion on the surface; complexation; catalysis; decomplexation; desorption.[34b]

A typical cellulase system comprises three main activities: chain-end-cleaving cellobiohydrolases (CBHs), internally chain cleaving endoglucanases (EGs), and ß-glucosidases (BGLs), which hydrolyze soluble short-chain glucooligosaccharides to glucose. Furthermore, CBH enzymes exhibit a preference for hydrolysis of the cellulose chain from either the reducing (CBH I) or the non-reducing end (CBH II). The three major hydrolytic enzymes act synergistically.[3,34]

Recent studies with modern methods, such as protein structural analysis,[40] in situ atomic force microscopy (AFM),[34a,41] PALM with two-dimensional total internal reflection fluorescence (TIRF–PALM), the investigation pf cohesin–dockerin interactions,[30] and differential kinetic analysis,[34b] elucidated details of the complex interplay of enzymes with the substrate.[3] In situ imaging (e.g. by AFM) offers insight into dynamic features of enzyme action and cellulose deconstruction down to the mesoscopic and nanoscale morphological level, and even down to the level of single enzyme molecules. This insight into cellulase activity on the cellulose surface enables quantitative information on surface dynamics to be obtained and obstacles identified, including rate-determining factors. The various visualization methods employed in different studies have substantially advanced our knowledge and helped to build and refine the model of enzymatic cellulose degradation.[34a]

The adsorption of different cellulases is mediated by CBMs, two common types of which (with specificity for either crystalline or amorphous regions) have been identified on cellulolytic enzymes. Crystalline-specific CBMs tend to cluster on ridges, which are typical of linear regions with glucan chains oriented parallel to one another. *Trichoderma*



Scheme 3. Hydrogen-bonding patterns in cellulose Iα and Iβ according to Sturcova et al.[35] a,b) The two alternative hydrogen-bond networks in cellulose Iα. c,d) The dominant hydrogen-bond network in cellulose Iβ chains at the origin of the unit cell (c) and in cellulose Iβ chains at the center of the unit cell (d).

reesei cellobiohydrolase I (TrCBH I) binds preferentially to the hydrophobic (110 or "planar") face of crystalline cellulose microfibrils.[42] CBH I binding sites on crystalline cellulose are limited owing to the tight packing of the chains and the likely burial of chain ends in inaccessible regions within the crystal. Amorphous-specific CBMs bind more homogeneously across the surface; they bind tightly to exposed glycan chains.[34b] EGs show a preference for regions thought to be amorphous.[3,34a] CBMs may also assist in nonhydrolytic substrate disruption (amorphogenesis).[43] Furthermore, other nonhydrolytic proteins have been demonstrated to aid in amorphogenesis.[44]

Diffusion of the enzymes into (voids of) the solid substrate was investigated and modeled by Luterbacher et al.,[38] but is not treated herein for two reasons: First, it depends on particle size, which may be adjusted to an appropriate size by pretreatment, and second, the reaction rates decrease dramatically after the initial burst, thus reducing the influence of diffusion.

When studying *Trichoderma reesei* TrCBH II on cellulose (Valonia) crystals, Chanzy and Henrissat[45] observed that the enzyme acts from one distinct end, the nonreducing end of the elongated crystals, thereby sharpening the tip of one end of the crystal. This tip-sharpening effect suggests that the degradation of ordered cellulose proceeds layerwise, starting from the readily accessible outer cellulose chains at a crystal edge. In contrast, TrCBH I starts degradation from the reducing end.[46] Additional evidence was presented of a TrCBH I enzyme sliding along a microfibril,[42b] and in high-speed AFM studies, Igarashi et al.[41] even determined the velocity of TrCBH I to be 3.5 $nm\,s^{-1}$.

3.5.1. Synergism

The cellulase complex with different functions and activities acts synergistically (and it must do so for efficiency). Cooperativity and synergism are not fully understood.[3] Synergism means that the hydrolysis rate of a combination of the three essential enzymes is significantly higher for cellulose degradation than the sum of the rates of the single enzymes.[3,34,47] Synergistic effects of a combination of endo- and exoenzymes were observed early with each of the CBH enzymes.[45,48] Endo–exo synergism is generally explained by the generation of new chain ends for CBH attack by a randomly cutting EG.[49] In more detail, EG and CBH II make the cellulose surface accessible for CBH I by removing amorphous unordered substrate areas, thus exposing otherwise embedded crystalline ordered nanofibrils of the cellulose. Subsequently, these fibrils are degraded efficiently by CBH I, thereby uncovering new amorphous areas. Thus, it has been concluded that synergism among cellulases is morphology-dependent and governed by the cooperativity of these enzymes.[34a]

Biochemical studies showed that synergistic degradation of the polymorphic substrate by the complete cellulase system resulted in 90% conversion; the rate of sugar release was enhanced fourfold as a consequence of the synergy between EG and CBH II. Furthermore, a sixfold increase in the hydrolysis rate after the addition of CBH I was observed. Hence, cooperativity between an endo–exo system (EG and CBH II), which prepares the cellulose, and an exocellulase (CBH I, with a pronounced preference for regions of higher order) is pivotal, and this morphology-dependent synergy therefore constitutes the primary synergism.[34a]

In an excellent study,[34a] the research group of Nidetzky used AFM to analyze the vertical degradation of polymorphic cellulose quantitatively on the basis of the complete cellulase system of *T. reesei* by evaluating multiple points on the surface and tracking their change over time with reference to large crystallites. The substrate was mixed amorphous–crystalline cellulose consisting of crystallites and small crystalline fibrils. The bulk of the amorphous material was degraded first. The large crystallites were degraded extremely slowly. The analysis led to the conclusion that cellulases remove amorphous material surrounding the crystalline fibrils first, and then degrade the fibrils to uncover subjacent amorphous material again. Fibrils are degraded in various ways (Figure 5), primarily by thinning starting from the side walls, but also by shortening starting from fibril tips and the introduction of defects in the middle of the fibrils (Figure 5c,d). As soon as a new defect was generated (Figure 5c, circle in the left panel), degradation proceeded rapidly from the two new ends thus produced (Figure 5c, arrows). Remarkably, distinct degradation velocities could be identified: When bared, small crystalline fibrils are degraded significantly faster ($(3.8 \pm 0.2)\ nm\,min^{-1}$) than the residual amorphous matrix ($(0.7 \pm 0.2)\ nm\,min^{-1}$). The reaction rate is strongly dependent on crystal size and shape and the accessibility of binding sites at crystals.[34a]

3.5.2. Reaction Kinetics

Kinetics can reveal key information about the mechanistic details of a reaction, as well as for reaction engineering, reactor design, and scale-up. However, no simple, mechanistically derived kinetics can be applied owing to the complexity of the substrates, with hemicellulose and lignin components, the use of pretreatment methods, and the range of complex events and sequences (see above). Therefore, no consistent rational model that includes all effects and findings is available.[50] A review by Bansal et al.[51] provides an overview of reported concepts and models of cellulose kinetics, and hence only a short summary is given herein. A mathematical model—based on experimental data—should comprise the most relevant enzyme activities (endo-acting EG, exo-acting CBH, BGL) to provide a rational basis for the optimal design of the most appropriate enzyme mixture, which will depend on the structure of the substrate and the pretreatment method applied (also xylanases, arabanases, galactanases, etc. should potentially be considered).

Endoglucanases were shown to potentiate the activity of cellobiohydrolases, not only by generating free chain ends but also by altering the substrate surface in such a way that the processive length of the enzyme is increased. Possible reaction steps discussed are (Figure 6; with reference, for example, to Ref. [41,47,52]): 1) A cellobiohydrolase is adsorbed to a crystalline region of the cellulose surface through its CBM domain; 2) it diffuses along the surface and forms

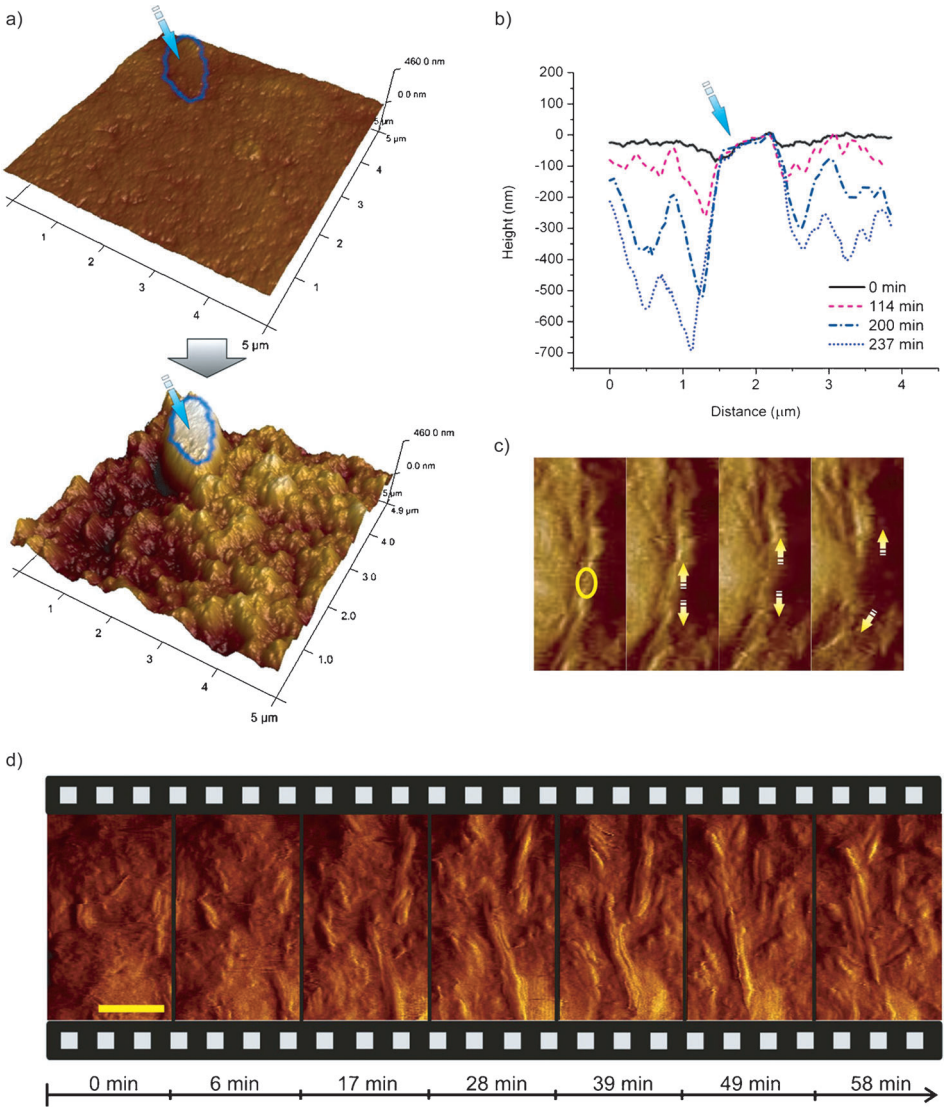


Figure 5. In situ observation of the synergistic degradation of polymorphic cellulose by the complete cellulase system of *T. reesei*. a) Substrate surface at the beginning of ($t = 0$ min) and after incubation ($t = 237$ min) with the complete cellulase system. A typical large crystallite, which was used as a marker for height changes during degradation, is indicated. The substrate around it was degraded, whereas the crystallite itself remained virtually unaltered. b) Section profiles of the area shown in (a) reflect the volumetric degradation with time. Relative to the marker, the substrate around it was degraded by up to 700 nm in 237 min. c) In situ observation of a defect (circle) introduced into a fibril. Subsequently, the fibril is attacked and degraded from both new ends generated by the defect (arrows). d) Snapshots along the time course of cellulose degradation show initial degradation of amorphous regions (17 min). Previously buried crystalline fibrils appear (28 min) and are quickly degraded in a processive manner starting at their ends as well as at previously introduced defects (49 min). Scale bar: 100 nm.[34a]

complexes with a cellulose chain end through its catalytic domain; 3) it catalyzes an initial hydrolysis event to generate glucose, cellobiose, or cellotriose (G1, G2, or G3, respectively); 4) it catalyzes subsequent hydrolysis events, generating only G2, as it moves processively along the chain; 5) it decomplexes from the cellulose chain; 6) it recomplexes with a cellulose chain or desorbs from the surface. Endoglucanases undergo similar steps, but they are generally thought to engage in nonprocessive random attack of amorphous or less crystalline regions with exposed single chains.

Kinetic data for individual steps of a model system (hydrolysis of bacterial microcrystalline cellulose, BMCC, by a cellobiohydrolase alone and in the presence of an endoglucanase) were analyzed quantitatively by Fox et al.[53] by the use of differential equations. The analysis revealed that the rate of the *Trichoderma longibrachiatum* cellobiohydrolase I (*Tl*Cel7A)-catalyzed hydrolysis of crystalline cellulose is limited by the rate of enzyme complexation with glycan chains, which was shown to be equivalent to the rate of generation of the initial-cut product (G1, G2, G3). This rate is enhanced in the presence of endoglucanase enzymes.[53] The results of the study and previous investigations show that the *Tl*Cel7A cellobiohydrolase enzymes engage in unidirectional processive catalysis on the hydrophobic face of the crystal,[41,42b] and that cellobiohydrolase complexation is slow relative to other kinetic steps and is thus rate-limiting (Figure 5).[53] Estimates of processive length for different enzyme mixtures on the basis of the ratio of the concentrations of processive-cut products and initial-cut products for a processive enzyme gave values in the range of 10–20 single reaction steps. Surprisingly, these values are lower than estimates of the intrinsic processive length of the *Trichoderma* Cel7A enzyme on this substrate by two orders of magnitude—a contradiction that has not been resolved.[53] The authors concluded that the surface morphology limits processivity, and that processive cellobiohydrolases are probably hindered by physical obstructions or other features that favor decomplexation, consistent with previous studies showing that the processive length depends on the nature of the substrate.[52]

Advanced mechanistic models of the mechanisms of selected steps have been developed by the research groups of Clark and Blanch,[54] who used model substrates (microcrystalline cellulose, whereby the solid cellulose substrate was represented as an assembly of spherical particles or as BMCC) and single activities, as well as mixtures of endo-

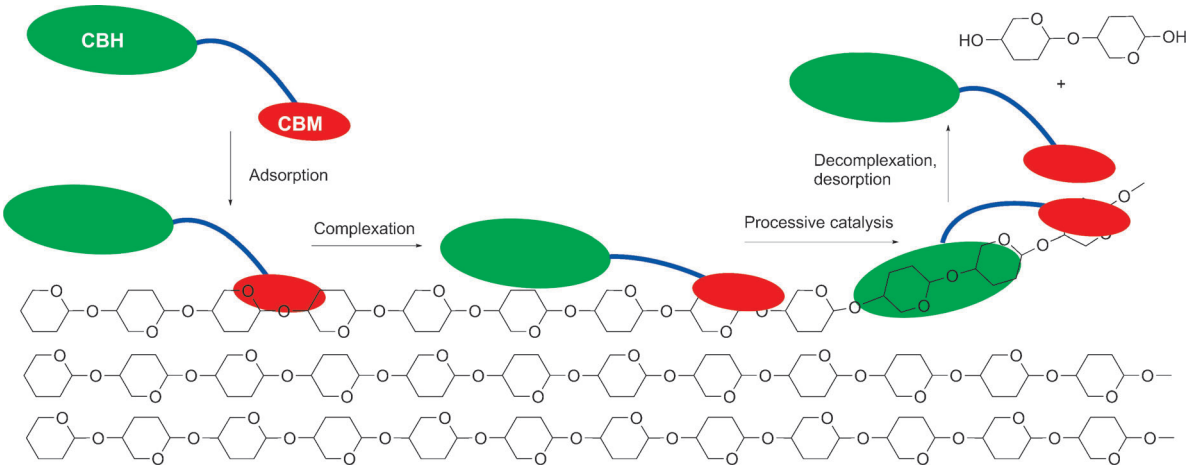


Figure 6. General reaction path used by cellobiohydrolases for the hydrolysis of insoluble cellulose (sugars are simplified). Reaction steps for strict endoglucanases are similar, but do not occur in processive catalytic sequences.[53]

and exocellulases. They included product inhibition, thermal deactivation, and surface heterogeneities of the substrate. However, they did not distinguish amorphous and crystalline parts, nor did they consider real, more complex substrates or β-glucosidase activity. Nevertheless, the results provide detailed insight into individual reaction steps. The article critically reviews previous kinetic models, the utility of which has been severely limited by further simplifying assumptions.

Results obtained for individual enzymes (*T. reesei* endoglucanase 2, EG2, and cellobiohydrolase I, CBHI) and systems with both enzymes present were compared with previously reported experimental data. The model was sensitive to cellulase-accessible surface area; the EG2–CBHI synergy observed experimentally was only predicted at a sufficiently high cellulose surface area, and the importance of separating the enzyme adsorption and complexation steps was shown by the model; the individual cellulose chain length and the time course with respect to the cellulose surface area (shrinking) were also taken into account. The adsorption and desorption of the cellulases were described according to the balance between binding sites and enzymes; this balance provides the concentration of cellulose surface sites and solution-phase enzymes. Balance and differential equations for the enzymes (including soluble, adsorbed, and inhibited (not shown here)) and the surface concentration of solid cellulose chains (length $i > 6$) were derived [Eq. (1)[*]]. A differential equation for the formation of cellobiose as the major soluble product is given in Equation (2).[54a] Kinetic parameters were taken from the literature.

$$\frac{\mathrm{d}[C'_i]}{\mathrm{d}t} = \frac{k_{\mathrm{cat-EG-cel}}}{k_{\mathrm{M-EG-cel}}}[\mathrm{E'}_{\mathrm{EG-ads}}]\theta_{\mathrm{EG}}\left(2\sum_{j=i+1}^{\infty}[C'_j] - (i-1)[C'_i]\right) + \frac{k_{\mathrm{cat-CBH-cel}}}{k_{\mathrm{M-CBH-cel}}}[\mathrm{E'}_{\mathrm{CBH-ads}}]\theta_{\mathrm{CBH}}([C'_{i+2}] - [C'_i]) \quad (1)$$

$$\frac{\mathrm{d}[C_2]}{\mathrm{d}t} = \left(2\frac{k_{\mathrm{cat-EG-cel}}}{K_{\mathrm{M-EG-cel}}}[\mathrm{E'}_{\mathrm{EG-ads}}]\theta_{\mathrm{EG}}\sum_{i=7}^{\infty}[C'_i] + \frac{k_{\mathrm{cat-CBH-cel}}}{K_{\mathrm{M-CBH-cel}}}[\mathrm{E'}_{\mathrm{CBH-ads}}]\theta_{\mathrm{CBH}}\sum_{i=7}^{\infty}[C'_i]\right)\frac{A_{\mathrm{cel}}}{V_{\mathrm{liq}}} \quad (2)$$

With significant enhancement of cellobiose-inhibition and thermal-inactivation parameters, the model was able to capture the slowdown in the hydrolysis rate exhibited by the experimental data. Strong cellobiose inhibition was required: $K_i = 0.01$ mM for EG2 and $K_i = 0.093$ mM for CBHI. The inclusion of short thermal-inactivation half-lives made quantitative agreement of the model with the experimental data possible, including EG2–CBHI synergy. These half-lives, however, were up to an order of magnitude smaller than those reported for a *T. reesei* cellulase mixture. Additional mechanisms that limit enzyme activity therefore need to be explored.[54a] (See also Ref. [53] for additional experimental data.)

The inhibition of cellobiohydrolase enzymes by glucose and cellobiose is well-known. In a recent study, Bu et al.[55] calculated the absolute free energies of binding of cellobiose to the Cel7A catalytic domain of *T. reesei*. For technical applications of the conversion of cellulosic material, experimental investigations remain mandatory with respect to (pretreated) substrates and their concentration, different enzyme mixtures and their concentration, reaction time, the pH value, and temperature for optimal results in terms of minimizing the amount of enzyme required, high conversion and yield, and high end-product concentration.[2a] For example, experiments with Celluclast showed the feasibility of designing minimal enzyme mixtures for pretreated lignocellulosic biomass with different compositions to that of the original enzyme mixture. Thus, enzyme mixtures with statistically designed combinations of the four main activities (two CBHs and two EGs) were used.[56] A large number of empirical models have been reported that help in the

[*] E′ stands for uncomplexed adsorbed enzyme, and θ is the fraction of free sites on the surface available to either CBH or EG; subscripts: ads stands for adsorbed enzyme, cel for solid cellulose, and *i*, *j* for the length of the cellulose chain. Equation (2) is simplified to show the principle of the kinetics. It does not contain terms for soluble intermediates with 3–6 glucose units, which are assumed to have a low concentration. For the full equations, see the original publication.[54a]

quantification of various effects observed. South et al. expressed the reaction rate constant in terms of conversion [Eq. (3)]:[57]

$$k(x) = k(1-x)^n + c \quad (3)$$

In this equation, k is the reaction rate constant for hydrolysis, x is conversion, $k(x)$ is the reaction rate constant at conversion x, n is the exponent of the declining rate constant, and c is a constant. The parameters n and c were estimated by approximating $k(x)$ by fitting it to the conversion (x). This expression was later used in modeling simultaneous saccharification and fermentation (SSF) with staged reactors and intermediate feeding of enzymes and the substrate.[58] Adsorption equations, for example, the Langmuir isotherm, have been incorporated into hydrolysis models, whereby the conversion range and four different cellulosic substrates were taken into account.[47,51]

Another empirical model for a commercial cellulase complex (Cellubrix, Novozymes) and both a model substrate (Avicel) and pretreated wheat straw was developed by Drissen et al.[59] Phenomena included in the mathematical model were the adsorption of the enzyme onto the substrate, substrate heterogeneity, including the effect of substrate conversion (termed "substrate recalcitrance"), glucose inhibition, the temperature dependency of reaction rates (with the activation energy $E_a = 29.8\ \mathrm{kJ\,mol^{-1}}$), and thermal enzyme inactivation (assuming a first-order type of process, $E_a = 148\ \mathrm{kJ\,mol^{-1}}$).

4. Engineering Aspects and Pilot and Demonstration Plants

Processing and reaction engineering must provide reactor concepts for the residence time required to obtain both high conversion and high product-concentration levels. Most importantly, feedback inhibition must be minimized to minimize the amount of enzymes required. The process steps of pretreatment, hydrolysis, and fermentation need to be viewed holistically to maximize ethanol yield and the overall cost of the process. Starting in 2000, Genencor and Novozymes successfully carried out major research programs, with considerable funds from the US Department of Energy (DOE), to reduce the cost of making cellulases from more than US$ 1.3 to less than 5.3 cents per liter. The enzymatic hydrolysis can either be performed separately from fermentation (SHF, separate hydrolysis and fermentation) or in combination with fermentation (SSF, simultaneous saccharification and fermentation). SSF avoids the primary drawback of separate hydrolysis caused by product inhibition of the enzymes by the released mono- and disaccharides. Compromises in the reaction temperature and pH value must be made in the design of the process: Hydrolysis by *T. reesei* enzymes is optimally performed at pH 5 and 50 °C, but adjustment to fermentation conditions at pH 5.5–7 and 30–40 °C is required.[60] In a model calculation for a SSF process by Galbe and Zacchi in 2007, the overall ethanol yield was 296 L per ton of dry feedstock, including ethanol losses in the process; the fact that part of the sugars are used for yeast production was also taken into consideration.[60d] This yield corresponds to 69 % of the theoretical value based on the hexose content in the raw material.

However, new developments in enzyme varieties tend to favor SHF for several reasons: A new enzyme cocktail (Cellic CTec 2, Novozymes) gives a 16–17 % higher final ethanol yield when applied separately in the hydrolysis, with an optimal (higher) temperature, with higher β-glucosidase activity, which reduces inhibition by cellobiose, and with an additional activity, a lytic polysaccharide monoxygenase (LPMO), which requires oxygen.[61] The investigations were undertaken with wheat straw, which was submitted to hydrothermal pretreatment (at 195 °C with a residence time of 18.5 min, followed by a washing step). Operations were conducted at very high dry-matter (DM) content (20 and 30 % DM), which is considered essential to make bioethanol production feasible in economic terms. Moreover, the surfactant PEG3000 was added in a further set of experiments at 30 % DM to investigate the possible beneficial effect on final yields under industrially relevant conditions; it is assumed that the surfactant prevents nonproductive enzyme adsorption to lignin. For SHF, the hydrolysis was carried out at 50 °C for 72 h, and fermentation at 34 °C. To improve mixing, which is inefficient in conventional reactors owing to the very high dry-matter content, a system based on roller bottles was applied. High glucose concentrations (maximum 150 $\mathrm{g\,kg^{-1}}$) were attained. The best results (summarized in Tables I and II of Ref. [61]) were obtained with the new enzyme mixture (dosage 22.8 mg per gram of cellulose; Cellic CTec 2, Novozymes): 1) SHF: 88.8 % ethanol yield (with reference to cellulose), hydrolysis at 50 °C, 72 h, 30 % DM, addition of PEG (0.4 %; 30 % saving of enzymes), fermentation at 34 °C, 96 h; 2) SSF: 86.3 % ethanol yield, same enzyme dosage, processing at 34 °C, 168 h, 20 % DM.

A range of pilot studies and demonstration and/or commercial production facilities have been announced in recent years; however, only a few have been in operation on a relevant scale, for example, by POET (USA), Iogen Corporation (Canada), Renmatrix, Inbicon in Denmark, and Abengoa in Spain (see also http://www.poet.com).[60b,61] Start-up facilities, mostly in the USA, are expected to be completed in 2012 or 2013 by Poet DSM Advanced Biofuels, KiOR, Beta Renewables, and ZeaChem, who will use, for example, corn stover or wheat straw as raw materials; GraalBio of Brazil will use sugarcane bagasse.[62]

A survey of 11 major projects was given by McCoy in 2012, with annual capacities up to 100 000 $\mathrm{m^3}$.[63] Iogen Corporation has operated the first demonstration facility in Ottawa (Canada) since 2004, in which a combination of thermal, chemical, and biochemical techniques are used to convert biomass as well as wheat, oat, and barley straw into ethanol. A modified steam explosion process is used for pretreatment. For enzymatic hydrolysis, Iogen developed a multistage system followed by fermentation. Efficient cellulase enzyme systems are produced in-house by the use of *T. reesei*. Fermentation is performed by microorganisms that convert both C_6 and C_5 sugars into ethanol. The yield is 340 L per ton of fiber (wheat straw). Large-scale-process

design includes efficient heat integration, water recycling, and coproduct utilization for overall process economy.[64] A demonstration plant in which straw and enzymes from Danisco Genencor, Novozymes, and DSM are used is operated in Kalundborg (Denmark) and produces 5400 m^3 ethanol per year, along with lignin pellets and molasses as by-products.[65] The plant of Chemtex in Italy, as well as the commercial-scale biorefinery of Beta Renewables in Crescentino, with a capacity of 75000 m^3 per year, both run on wheat straw and a giant reed, *Arundo donax*, with enzymes from Novozymes. Beta Renewables indicates a sugar price from biomass of 22 cents per kg, as compared to 40–44 cents per kg for sugar from corn, and a price for ethanol of 65 cents per liter.[62a] BASF (Germany) will work with Renmatrix (USA) to scale up a process for producing sugars from, for example, wood or straw, by using water at high temperature and pressure; the sugars then can be fermented to produce intermediates such as 1,4-butanediol and succinic acid.[66] Clariant developed the sunliquid process, for which no additional fossil energy is required, for the synthesis of ethanol from lignocellulosic plant material. A pilot plant for this process went into operation in 2012 in Straubing (Germany) with an annual capacity of up to 1000 tons of ethanol.[67]

Plant genetic engineering may open new perspectives for the more straightforward hydrolysis of lignocellulosic material. One approach is to change the lignin content and structure, as has been shown recently for mutants of *Arabidopsis thaliana* with respect to an enzyme central to the lignin biosynthetic pathway.[68] These mutants deposited less lignin than do wild-type plants. As a result, the conversion of cellulose into glucose in the mutants was up to fourfold higher than in the wild type upon saccharification without pretreatment.

5. Discovery of Novel CAZymes

The traditional method for the identification of new enzymes is based on the screening, for example, of soil samples, unusual habitats, or industrial sites, of strain collections by enrichment culture, or of plant or animal tissues. This general concept has also been applied for the discovery of CAZymes, but this approach will miss the majority of the biodiversity of nature, as only a small number of microbes can successfully be cultivated in the laboratory. As pointed out by Wilson in an excellent review, two major problems limit our understanding of the degradation of cellulosic material: "*the vast diversity of organisms present in most cellulose degrading environments and the inability to culture most of them.*"[69] Depending on the habitat, it was estimated that less than 1% of the microbes present in an environmental sample can be grown under standard laboratory conditions.[70] Reasons are that appropriate cultivation conditions are unknown, growth rates are too slow, or the growth of a microbe depends on other species, which provide nutrients. The metagenome approach overcomes these limitations by circumventing the cultivation step. In this method, the complete genomic DNA is extracted from a sample, fragmented, and cloned to yield the corresponding metagenome libraries.[71] These libraries can then be screened either for the desired activity (with a function-based assay) or in a sequence-based approach, in which the entire DNA is sequenced and then genes are identified on the basis of homology to already described enzymes. Recent reviews nicely summarize enzymes, including CAZymes, discovered by both approaches.[72] These reviews also cover studies on the discovery of novel glycoside hydrolases from ecosystems such as soil, lakes, hot springs, rumina, and rabbit and insect guts.

In an outstanding effort, the human gut microbiome was explored by a metagenomic approach to discover new CAZymes.[73] The reasoning behind the exploration of the microbial community present in the human intestine is that dietary fibers are mainly composed of complex cell-wall polysaccharides and lignins. Because of the wide structural diversity of dietary fibers, the authors assumed that human gut bacteria—of which only 20% could be successfully cultured by using standard microbiology methods—must produce a huge panel of CAZymes. In earlier studies, this biodiversity was already investigated by functional and structural genomics analysis restricted to culturable bacteria.[74]

Tasse et al. first created and screened a metagenomic library (5.4×10^9 base pairs of metagenomic DNA cloned and expressed in *Escherichia coli*) by using functional plate-based assays for β-glucan, xylan, galactan, pectin, and amylose degradation for a total of 704000 clones.[73] Subsequently, 310 clones were subjected to a secondary screen with the aim of deciphering enzyme specificity towards 15 polysaccharides and enzyme stability at different pH values and temperatures. Active clones were pyrosequenced, followed by gene annotation and taxonomic assignation. This effort resulted in the discovery of 18 CAZy gene clusters comprising 73 CAZy genes from 35 families and 86 CAZy modules. Importantly, many novel genes discovered had no high sequence homology to known proteins, which confirms that this strategy allows the identification of completely novel enzymes. In a more recent study, the same group extended this concept to two metagenomic libraries from ileum mucosa and fecal microbiota with a focus on the degradation of probiotic carbohydrates, such as inulin and fructooligosaccharides.[75] Again, a broad set of novel carbohydrate-hydrolyzing enzymes belonging to various CAZyme families were identified, as well as transporters. Besides the use of agar-plate assays and standard analytical methods, an NMR-based tool for the high-throughput screening of CAZymes was also developed to enable the quantitative identification of reaction products and provide structural information.[76] The method involving gene-library expression and direct analysis of the crude reaction medium was automated for carbohydrate-structure determination. Thus, a library of 4032 CAZymes could be analyzed at a rate of 480 enzyme variants per day; 303 of these enzymes could be accurately characterized.

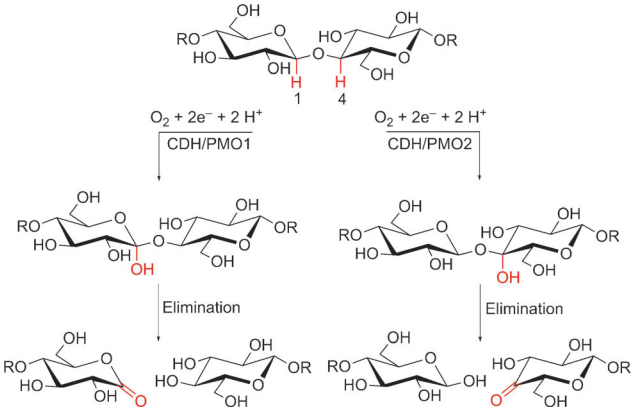
In another outstanding study, the metagenome approach was used to discover cellulosic-biomass-degrading enzymes in the cow rumen with switchgrass as the target substrate.[77] The authors sequenced and annotated 268 gigabases of DNA from microbes present in cow rumen. In this way, they identified

27755 genes encoding for putative carbohydrate-active proteins. From 90 candidates expressed, 57% were confirmed active against a panel of ten cellulosic substrates. An important aspect of this study is that the in-depth sequencing and comparison of the gene sequences discovered with the entries in the CAZyme database revealed that only 1% of the 2716 sequences analyzed (encoding for β-(1,4)-endoglucanase, β-glucosidase, or cellobiohydrolase) were highly similar to any CAZy database entry. In a recent study, oxygen-tolerant microbial consortia capable of the decomposition of lignocelluloses in agricultural residues and industrial pulp waste were analyzed.[78] The authors found a stable coexistence of *Clostridium*, *Acetanaerobacterium*, and *Ureibacillus* species with CMCase, xylanase, and β-glucanase activities. Shotgun pyrosequencing of one metagenome library identified a variety of CAZymes belonging to 26 families. These examples strongly indicate that the vast majority of CAZymes have escaped identification by standard methods, and that only metagenomic approaches combined with extensive sequencing allow their discovery.

An almost untapped source of potential CAZymes are Archea.[79] In one example, *Desulfurococcus fermentans* was identified as a hyperthermophilic organism and found to be able to grow on crystalline cellulose at 81 °C. In a more recent study, Graham et al.[80] found that only a consortium of three hyperthermophilic archaea obtained from a continental geothermal source was able to act on crystalline cellulose by growth at 90 °C. Interestingly, attempts to separate the individual species failed. It was argued that a single isolated hyperthermophilic archaeon is unable to produce the various enzymes needed for the complex degradation of lignocellulosic substrates, possibly because of the restricted size of their individual genomes. Subsequent metagenomic analysis of the consortium identified a 90 kDa multidomain endoglucanase with uncommon structural features. This enzyme has astonishing properties, with optical activity at 109 °C and a half-life of 5 h at 100 °C, and was found to be very stable in the presence of ionic liquids, strong detergents, and high salt concentrations.

Recently, the importance of oxidative enzymes for the degradation of (ligno-)cellulosic material was emphasized, and many known enzymes together with newly discovered activities were reclassified as "auxiliary enzymes" in the CAZy database.[23,81] This subfamily now includes well-studied enzymes, such as laccases, manganese peroxidases, and lignin peroxidases, all of which are multicopper oxidases known to be essential for oxidative lignin conversion. Other "auxiliary enzymes" are cellobiose dehydrogenases, pyranose oxidase, and galactose oxidase. Cellulases so far classified as GH61 enzymes were recently shown to be copper-dependent (lytic) polysaccharide monooxygenases (LPMOs) that enhance cellulase degradation in concert with classical cellulases (see Ref. [82] and references cited therein). These fungal-derived PMOs from organisms such as *Neurospora crassa* catalyze the oxidative cleavage of cellulose by using reducing agents, such as ascorbate, gallate, or reduced glutathione. Most importantly, PMOs can act directly on crystalline microfibril cellulose. Quinlan et al. reported a type II copper site as the metal involved in this newly observed GH61 activity.[83] Marletta and co-workers first identified the oxidation products at the nonreducing end as 4-ketoaldoses and could thus confirm that no C6 oxidation took place. By isotope labeling, they proposed a mechanism—depending on whether a type I or a type II PMO (PMO1 or PMO2) is used—by which oxygen insertion occurs at the C1 or C4 position, followed by elimination leading to the ketoaldoses (Scheme 4).[84]

Very recently, another family of auxiliary enzymes (AA11) was described,[85] thus increasing the number of known LPMOs that act on cellulose (AA9, formerly GH61) or chitin (AA10, formerly CBM33). The AA11 enzyme from *Aspergillus oryzae* was shown to have oxidative chitinolytic activity, which yields aldonic acid oligosaccharides. Biochemical studies revealed that the properties of the AA11 enzymes are structurally and spectroscopically in between those of the AA9 and AA10 families.[82b]



Scheme 4. Reaction pathway proposed for the oxidative cleavage of cellulose by PMO1 or PMO2.[83]

CBM33 modules were reclassified and also resemble copper-dependent LPMOs. In a pioneering study, CBP21 was first shown to be in fact an oxidative enzyme that catalyzes oxidative cleavage of crystalline chitin, which is boosted by the addition of ascorbate to yield the oxidized sugar 2-(acetylamino)-2-deoxy-D-gluconic acid.[86] Chauvigné-Hines et al. developed probes with alkyne tags for activity-based protein profiling (ABPP) for the direct labeling of the *C. thermocellum* cellulosomal secretome.[87] ABP selectivity for glycoside hydrolase (GH) enzymes was followed by mass spectroscopy.

6. Protein Engineering of CAZymes

Protein engineering is nowadays a well-established technology for altering the properties of enzymes, and general strategies, as well as numerous successful examples, are well-documented.[88] The two principle concepts are rational protein design or directed (molecular) evolution. If the 3D structure or a reasonable homology model is available, rational protein design is the method of choice to predict the type and position of amino acid mutations, which are then

Table 1: Selected examples of the protein engineering of CAZYmes.[a]

Enzyme family	Source organism	Mutation method	Assay/ Substrate	Improved property	Remarks	Ref.
Endoglucanase (endo-β-1,4-glucanase), EC 3.2.1.4						
GH5/Cel5A	*Bacillus subtilis*	epPCR/DNA shuffling	Congo red staining/ DNS/ CMC	increased activity (up to 2.7-fold) towards CMC	slightly improved pH tolerance and thermostability; up to 12 mutations, most outside conserved regions	[96]
GH5/Cel5A	*Clostridium phytofermentans*	epPCR	Congo red staining, CMC	slightly improved half-lives for CMC, amorphous cellulose, and Avicel	surface display of a fusion protein; expression with or without CBM or a His tag; strong variation of mutant properties as influenced by these constructs	[93]
GH5/CBM3	*Bacillus amyloliquefaciens*	error-prone rolling circle amplification	DNS/ CMC	increased activity (up to 8-fold) towards CMC	six mutations found, but only E289 V relevant, as situated in the cellulase catalytic domain	[97]
Cel8A	*Clostridium thermocellum*	epPCR and consensus mutations	PASC/ CMC	improved thermostability with a 14-fold longer half-life at 85 °C	consensus approach identified a G283P mutation, which was combined with a previous triple mutant	[98]
EGII	*Trichoderma reesei*		Congo red staining/ Avicel/ CMC	enhanced cellulase activity by improved binding of CBM	CBM library targeting two residues created by cell-surface engineering in yeast; 1.5-fold higher binding ability and 1.3-fold higher hydrolytic activity	[99]
CMCase	*Bacillus subtilis*	shuffling	Congo red staining/ CMC, HPLC	variants with 2–5-fold improved activity towards CMC	bacterial surface display for library screening; up to 8 mutations, most in the catalytic domain	[100]
Cellobiohydrolase, EC 3.2.1.91						
GH6/Cel6A/ CBH II	three homologous enzymes, e.g. from *Humicola insolens*	SCHEMA recombination of 8 blocks	PASC/ Avicel	improved thermostability at 63 °C; variants active at temperatures 7–15 °C higher than for the wild type	15 variants with up to 63 mutations investigated in detail, also variants with an improved pH profile; a later study showed that a single mutation was responsible	[101]
GH6/Cel6A	best variant from a previous study[100]	epPCR/recombination	Avicel	half-life of 280 min at 75 °C, T_{50} = 80 °C (15–20 °C higher than for the wildtype)	crystal structures of mutants reported	[102]
GH6/CBH II	*Phanerochaete chrysosporium*	consensus approach	PASC	modestly improved thermostability at 50 °C for 72 h	IVTT for enzyme production; best variant contained 15 mutations, which resulted in an improved ΔT_m value by ca. 5 °C	[103]
GH7/CBH I/ Cel7B	*Melanocarpus albomyces*	point mutations/fusion with CBM	MUL/ Avicel	thermoactivity and stability improved; addition of CBM caused a 7-fold activity increase towards Avicel	an additional disulfide bridge near the N terminus and a single mutation improved ΔT_m by 4 °C	[104]
GH7/CBH I/ Cel7A	*Talaromyces emersonii*	disulfide-bond engineering	MUL/ Avicel	thermoactivity and stability improved	five single S—S variants studied; one triple mutant had an improved ΔT_m value by 9 °C and activity at 80 °C	[105]
GH7/CBH I/ Cel7A	five homologous enzymes, e.g. from *Talaromyces emersonii*	SCHEMA recombination of 8 blocks	MUL/ solid cellulose	thermostability; chimeras active at 70 °C (wild type: 65 °C)	16 chimeras functionally expressed, on average 37 mutations per chimera	[106]

Table 1: (Continued)

Enzyme family	Source organism	Mutation method	Assay/ Substrate	Improved property	Remarks	Ref.
GH7/CBH I/ Cel7A	chimera derived from *Talaromyces emersonii*	rational design by FoldX/consensus concept	MUL/ solid cellulose	thermostability; stability increased by 10 °C to 65 °C, 50% increase in sugar release from CMC	computational analysis predicted 43 mutants, 9 were investigated, the best 8 combined; effective mutations were on protein surface	[107]
GH8/Cel7A	11 starting enzymes, e.g. from *Aspergillus* sp.	recombination	MUL/ Avicel	thermostability and activity; 86% of the 469 chimera were active, 51 had higher stability; on average 4–5 mutations per chimera	"biased clique shuffling" claimed to be superior to standard methods; also, a 100-fold better expression system was developed in *Saccharomyces cerevisiae* to avoid hyperglycosylation	[108]
Miscellaneous enzymes						
GH11/xylanase	not reported, metagenome origin	GSSM and recombination	oat-spelt xylan	substantially improved thermostability from 76 to 101 °C	most mutations in N-terminal region; 3D structure provided no clear hint for improved stability/activity and rather suggests subtle changes	[109]
GH70/reuteransucrase	*Lactobacillus reuteri*	rational design	growth assay	allosyltransferase activity	modeling predicted acceptance of novel donors	[110]
Cel48	recombination of three genes from *Clostridium* sp.	SCHEMA recombination	Avicel	large variations in temperature range, stability, and activity found	mutants contained up to 106 mutations in comparison to the wild types; 60 new Cel48 variants characterized	[111]

[a] DNS = dinitrosalicylic acid, CMC = carboxymethyl cellulose, CBM = cellulose-binding domain, MUL = 4-methylumbelliferyl β-D-lactoside, PSAC = phosphoric acid swollen cellulose, IVTT = in vitro transcription translation, GSSM = gene-site saturation mutagenesis.

introduced into the protein-encoding gene by site-directed mutagenesis. Once the variant has been produced by recombinant expression, the enzyme can be purified and characterized for the desired property. In the absence of a protein structure or sufficient information to guide protein engineering, directed evolution, in which mutations are introduced randomly (e.g. by error-prone PCR)[89] or homologous genes are recombined (e.g. by DNA shuffling),[90] is the method of choice. The key challenge in directed evolution is the availability of a suitable high-throughput screening method for the quick and reliable identification of desired variants from the huge mutant library created by random mutagenesis.

Protein engineering was also successfully used for CAZYmes, initially with a focus on the identification of the catalytic mechanism,[91] followed by more recent research to improve their properties.[91,92] The engineering of these enzymes is very challenging because the enzymatic hydrolysis is complicated, heterogeneous substrates are involved, and various components are required for enzymatic activity. Directed evolution has been used to create variants with enhanced activity on soluble substrates, improved thermostability, a different optimum pH value, or improved expression. Improved thermoactivity (> 50 °C) is an important target, as it results in reduced contamination and viscosity as well as increased hydrolytic activity. Improved stability also leads to reduced enzyme costs. An overview of selected examples is given in Table 1.

As early as 2010, Liu et al. stated that "*the development of high-throughput cellulase assays on solid cellulosic substrates is urgently needed*".[93] In most examples, variants were first screened either in agar-plate assays or in microtiter plates by using soluble chromophores, such as 4-methylumbelliferyl β-D-lactoside or phosphoric acid swollen cellulose, in combination with a dye to separate active from inactive clones. Only in a subsequent second assay were real cellulosic substrates, such as Avicel, used for the identification of truely improved variants. A further challenge is that most CAZymes are not expressed well in standard hosts used for directed evolution (e.g., *E. coli*, *S. cerevisiae*). As a consequence, the best variants might be missed owing to a lack of proper expression, or variants identified to exhibit improved properties might fail in later large-scale studies when industrial production hosts are used. In particular, CBH I cellobiohydrolases have proven difficult to engineer, in part as a result of their low expression level in heterologous hosts, but mainly because of their complex structure. CBH I enzymes usually contain 8–10 disulfide bridges and at least three glycosylation sites. Whereas in most directed-evolution experiments either error-prone PCR or shuffling were used, and often resulted in only moderate improvements (Table 1), Arnold and coworkers were more successful in using the SCHEMA algorithm to guide mutant design. This structure-guided recombination method functions in the absence of high-throughput screening methods and identifies blocks suitable

for the creation of chimeric enzymes even at sequence identities as low as 30%.[94]

7. Summary and Future Perspectives

The ability to obtain (cheap) fermentable sugars from biomass is important not only for the production of ethanol, but also for many other fermentative processes in bulk chemistry, as the production of ethanol has to compete with recently established economically viable alternatives based on shale-gas utilization.[95] Recent developments, such as the oxygenases, are currently calling into question many aspects of the classic mechanism and are somehow reshaping the concept that we had of the enzymatic degradation of cellulose. Different cellulases and oxygenases enable the degradation of a wider variety of surfaces. The resulting monosaccharides and oligosaccharides can be reassembled by engineered enzymes into novel tailor-made products. The biorefinery of the future may also allow the use of lignin and hemicelluloses as substrates for the production of bulk compounds and fine chemicals if the repertoire of nature can be extended to novel enzyme activities. In this regard, the implementation of pretreatment, biocatalyst design, and reaction engineering will open new avenues for the synthesis of biofuels and chemicals.

A DuPont Young Professor Award for J.S. is gratefully acknowledged.

Received: November 15, 2013
Published online: August 18, 2014

[1] L. Caspeta, J. Nielsen, *Nat. Biotechnol.* **2013**, *31*, 789 – 793.
[2] a) S. Merino, J. Cherry, *Biofuels*, *Vol. 108* (Ed.: L. Olsson), Springer, Berlin/Heidelberg, **2007**, pp. 95 – 120; b) E. Gnansounou, A. Dauriat, *Bioresour. Technol.* **2010**, *101*, 4980 – 4991.
[3] P. Bubner, H. Plank, B. Nidetzky, *Biotechnol. Bioeng.* **2013**, *110*, 1529 – 1549; b) A. J. Ragauskas, G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. A. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. A. Tuskan, C. E. Wyman, *Science* **2014**, DOI: 10.1126/science.1246843; c) X. Meng, A. J. Ragauskas, *Curr. Opin. Biotechnol.* **2014**, *27*, 150 – 158.
[4] J. Tolan, B. Foody, *Recent Progress in Bioconversion of Lignocellulosics*, *Vol. 65* (Eds.: G. T. Tsao, A. P. Brainard, H. R. Bungay, N. J. Cao, P. Cen, Z. Chen, J. Du, B. Foody, C. S. Gong, P. Hall, N. W. Y. Ho, D. C. Irwin, P. Iyer, T. W. Jeffries, C. M. Ladisch, M. R. Ladisch, Y. Y. Lee, N. S. Mosier, H. M. Mühlemann, M. Sedlak, N. Q. Shi, J. S. Tolan, R. W. Torget, D. B. Wilson, L. Xia), Springer, Berlin/Heidelberg, **1999**, pp. 41 – 67.
[5] a) H. Chanzy, B. Henrissat, *Carbohydr. Polym.* **1983**, *3*, 161 – 173; b) B. Henrissat, H. Driguez, C. Viet, M. Schulein, *Nat. Biotechnol.* **1985**, *3*, 722 – 726.
[6] a) M. R. Ladisch, K. Dyck, *Science* **1979**, *205*, 898 – 900; b) M. R. Ladisch, B. E. Dale, G. T. Tsao, *Biotechnol. Bioeng.* **1983**, *25*, 1 – 2.
[7] T. K. Ghose, *Pure Appl. Chem.* **1987**, *59*, 257 – 268.
[8] a) "Cellulases": K. Buchholz, P. Rapp, F. Zadrazil in *Methods of Enzymatic Analysis*, *Vol. 3* (Ed.: H. U. Bergmeyer) Verlag Chemie, Weinheim, **1984**, chapter 4, pp. 178 – 188; b) K. Buchholz, J. Puls, B. Gödelmann, H. V. Dietrichs, *Process Biochem.* **1981**, *16*, 37 – 43.
[9] International_Energy_Agency, *Key World Energy Statistics 2012* **2012**. http://www.iea.org/publications/freepublications/publication/kwes.pdf
[10] H. Kopetz, *Nature* **2013**, *494*, 29 – 31.
[11] M. Sette, R. Wechselberger, C. Crestini, *Chem. Eur. J.* **2011**, *17*, 9529 – 9535.
[12] a) P. Alvira, E. Tomás-Pejó, M. Ballesteros, M. J. Negro, *Bioresour. Technol.* **2010**, *101*, 4851 – 4861; b) A. T. W. M. Hendriks, G. Zeeman, *Bioresour. Technol.* **2009**, *100*, 10 – 18.
[13] a) G. Brodeur, E. Yau, K. Badal, J. Collier, K. B. Ramachandran, S. Ramakrishnan, *Enzyme Res.* **2011**, 787532; b) U. Holopainen-Mantila, K. Marjamaa, Z. Merali, A. Käsper, P. de Bot, A.-S. Jääskeläinen, K. Waldron, K. Kruus, T. Tamminen, *Bioresour. Technol.* **2013**, *138*, 156 – 162; c) S.-Y. Leu, J. Y. Zhu, *Bioenerg. Res.* **2013**, *6*, 405 – 415; d) B. C. Saha, T. Yoshida, M. A. Cotta, K. Sonomoto, *Ind. Crops Prod.* **2013**, *44*, 367 – 372; e) M. A. Lima, G. B. Lavorente, H. K. da Silva, J. Bragatto, C. A. Rezende, O. D. Bernardinelli, E. R. Deazevedo, L. D. Gomez, S. J. McQueen-Mason, C. A. Labate, I. Polikarpov, *Biotechnol. Biofuels* **2013**, *6*, 75; f) R. Ibbett, S. Gaddipati, S. Hill, G. Tucker, *Biotechnol. Biofuels* **2013**, *6*, 33; g) H. Li, Y. Pu, R. Kumar, A. J. Ragauskas, C. E. Wyman, *Biotechnol. Bioeng.* **2014**, *111*, 485 – 492; h) B. Yang, C. E. Wyman, *Biofuels Bioprod. Biorefin.* **2008**, *2*, 26 – 40; i) C. Zetzl, K. Gairola, C. Kirsch, L. Perez-Cantu, I. Smirnova, *Chem. Ing. Tech.* **2012**, *84*, 27 – 35.
[14] B. Madsen, E. K. Gamstedt, *Adv. Mater. Sci. Eng.* **2013**, *14*.
[15] X. Zhao, K. Cheng, D. Liu, *Appl. Microbiol. Biotechnol.* **2009**, *82*, 815 – 827.
[16] J. Viell, A. Harwardt, J. Seiler, W. Marquardt, *Bioresour. Technol.* **2013**, *150*, 89 – 97.
[17] T. vom Stein, P. M. Grande, H. Kayser, F. Sibilla, W. Leitner, P. Domínguez de María, *Green Chem.* **2011**, *13*, 1772 – 1777.
[18] B. Henrissat, *Biochem. J.* **1991**, *280*, 309 – 316.
[19] B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, *Nucleic Acids Res.* **2009**, *37*, D233 – 238.
[20] a) B. Henrissat, A. Bairoch, *Biochem. J.* **1993**, *293*, 781 – 788; b) B. Henrissat, A. Bairoch, *Biochem. J.* **1996**, *316*, 695 – 696; c) G. Davies, B. Henrissat, *Structure* **1995**, *3*, 853 – 859; d) B. Henrissat, G. Davies, *Curr. Opin. Struct. Biol.* **1997**, *7*, 637 – 644.
[21] a) J. A. Campbell, G. J. Davies, V. Bulone, B. Henrissat, *Biochem. J.* **1997**, *326*, 929 – 939; b) P. M. Coutinho, E. Deleury, G. J. Davies, B. Henrissat, *J. Mol. Biol.* **2003**, *328*, 307 – 317.
[22] V. Lombard, T. Bernard, C. Rancurel, H. Brumer, P. M. Coutinho, B. Henrissat, *Biochem. J.* **2010**, *432*, 437 – 444.
[23] A. Levasseur, E. Drula, V. Lombard, P. M. Coutinho, B. Henrissat, *Biotechnol. Biofuels* **2013**, *6*, 41.
[24] H. J. Gilbert, J. P. Knox, A. B. Boraston, *Curr. Opin. Struct. Biol.* **2013**, *23*, 669 – 677.
[25] A. B. Boraston, D. N. Bolam, H. J. Gilbert, G. J. Davies, *Biochem. J.* **2004**, *382*, 769 – 781.
[26] A. L. Creagh, E. Ong, E. Jervis, D. G. Kilburn, C. A. Haynes, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12229 – 12234.
[27] S. Najmudin, B. A. Pinheiro, J. A. Prates, H. J. Gilbert, M. J. Romao, C. M. Fontes, *J. Struct. Biol.* **2010**, *172*, 353 – 362.
[28] C. M. Payne, M. G. Resch, L. Chen, M. F. Crowley, M. E. Himmel, L. E. Taylor II, M. Sandgren, J. Stahlberg, I. Stals, Z. Tan, G. T. Beckham, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14646 – 14651.
[29] a) E. A. Bayer, R. Kenig, R. Lamed, *J. Bacteriol.* **1983**, *156*, 818 – 827; b) R. Lamed, E. Setter, E. A. Bayer, *J. Bacteriol.* **1983**, *156*, 828 – 836.
[30] C. M. G. A. Fontes, H. J. Gilbert, *Annu. Rev. Biochem.* **2010**, *79*, 655 – 681.

[31] a) A. L. Carvalho, F. M. Dias, J. A. Prates, T. Nagy, H. J. Gilbert, G. J. Davies, L. M. Ferreira, M. J. Romao, C. M. Fontes, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13809 – 13814; b) S. W. Stahl, M. A. Nash, D. B. Fried, M. Slutzki, Y. Barak, E. A. Bayer, H. E. Gaub, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 20431 – 20436; c) A. L. Carvalho, F. M. Dias, T. Nagy, J. A. Prates, M. R. Proctor, N. Smith, E. A. Bayer, G. J. Davies, L. M. Ferreira, M. J. Romao, C. M. Fontes, H. J. Gilbert, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 3089 – 3094.

[32] G. Gefen, M. Anbar, E. Morag, R. Lamed, E. A. Bayer, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10298 – 10303.

[33] Y. Mori, S. Ozasa, M. Kitaoka, S. Noda, T. Tanaka, H. Ichinose, N. Kamiya, *Chem. Commun.* **2013**, *49*, 6971 – 6973.

[34] a) T. Ganner, P. Bubner, M. Eibinger, C. Mayrhofer, H. Plank, B. Nidetzky, *J. Biol. Chem.* **2012**, *287*, 43215 – 43222; b) J. M. Fox, P. Jess, R. B. Jambusaria, G. M. Moo, J. Liphardt, D. S. Clark, H. W. Blanch, *Nat. Chem. Biol.* **2013**, *9*, 356 – 361.

[35] A. Sturcová, I. His, D. C. Apperley, J. Sugiyama, M. C. Jarvis, *Biomacromolecules* **2004**, *5*, 1333 – 1339.

[36] Y. Habibi, L. A. Lucia, O. J. Rojas, *Chem. Rev.* **2010**, *110*, 3479 – 3500.

[37] Y. H. Zhang, L. R. Lynd, *Biotechnol. Bioeng.* **2004**, *88*, 797 – 824.

[38] J. S. Luterbacher, L. P. Walker, J. M. Moran-Mirabal, *Biotechnol. Bioeng.* **2013**, *110*, 108 – 117.

[39] J. S. Luterbacher, J.-Y. Parlange, L. P. Walker, *Biotechnol. Bioeng.* **2013**, *110*, 127 – 136.

[40] J. L. Bras, A. Cartmell, A. L. Carvalho, G. Verze, E. A. Bayer, Y. Vazana, M. A. Correia, J. A. Prates, S. Ratnaparkhe, A. B. Boraston, M. J. Romao, C. M. Fontes, H. J. Gilbert, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5237 – 5242.

[41] K. Igarashi, A. Koivula, M. Wada, S. Kimura, M. Penttila, M. Samejima, *J. Biol. Chem.* **2009**, *284*, 36186 – 36190.

[42] a) H. Chanzy, B. Henrissat, R. Vuong, *FEBS Lett.* **1984**, *172*, 193 – 197; b) Y. S. Liu, J. O. Baker, Y. Zeng, M. E. Himmel, T. Haas, S. Y. Ding, *J. Biol. Chem.* **2011**, *286*, 11195 – 11201.

[43] V. Arantes, J. N. Saddler, *Biotechnol. Biofuels* **2010**, *3*, 4.

[44] a) G. Jäger, M. Girfoglio, F. Dollo, R. Rinaldi, H. Bongard, U. Commandeur, R. Fischer, A. C. Spiess, J. Büchs, *Biotechnol. Biofuels* **2011**, *4*, 33; b) H. J. Lee, S. Lee, H. J. Ko, K. H. Kim, I. G. Choi, *Mol. Cells* **2010**, *29*, 379 – 385; c) R. E. Quiroz-Castaneda, C. Martínez-Anaya, L. I. Cuervo-Soto, L. Segovia, J. L. Folch-Mallol, *Microb. Cell Fact.* **2011**, *10*, 8; d) M. Saloheimo, M. Paloheimo, S. Hakola, J. Pere, B. Swanson, E. Nyyssonen, A. Bhatia, M. Ward, M. Penttila, *Eur. J. Biochem.* **2002**, *269*, 4202 – 4211.

[45] H. Chanzy, B. Henrissat, *FEBS Lett.* **1985**, *184*, 285 – 288.

[46] B. K. Barr, Y. L. Hsieh, B. Ganem, D. B. Wilson, *Biochemistry* **1996**, *35*, 586 – 592.

[47] B. Nidetzky, W. Steiner, M. Hayn, M. Claeyssens, *Biochem. J.* **1994**, *298*, 705 – 710.

[48] A. R. White, R. M. Brown, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 1047 – 1051.

[49] a) J. Jalak, M. Kurasin, H. Teugjas, P. Valjamae, *J. Biol. Chem.* **2012**, *287*, 28802 – 28815; b) L. R. Lynd, P. J. Weimer, W. H. van Zyl, I. S. Pretorius, *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 506 – 577.

[50] E. Vlasenko, M. Schülein, J. Cherry, F. Xu, *Bioresour. Technol.* **2010**, *101*, 2405 – 2411.

[51] P. Bansal, M. Hall, M. J. Realff, J. H. Lee, A. S. Bommarius, *Biotechnol. Adv.* **2009**, *27*, 833 – 848.

[52] M. Kurasin, P. Valjamae, *J. Biol. Chem.* **2011**, *286*, 169 – 177.

[53] J. M. Fox, S. E. Levine, D. S. Clark, H. W. Blanch, *Biochemistry* **2012**, *51*, 442 – 452.

[54] a) S. E. Levine, J. M. Fox, H. W. Blanch, D. S. Clark, *Biotechnol. Bioeng.* **2010**, *107*, 37 – 51; b) S. E. Levine, J. M. Fox, D. S. Clark, H. W. Blanch, *Biotechnol. Bioeng.* **2011**, *108*, 2561 – 2570.

[55] L. Bu, M. R. Nimlos, M. R. Shirts, J. Stahlberg, M. E. Himmel, M. F. Crowley, G. T. Beckham, *J. Biol. Chem.* **2012**, *287*, 24807 – 24813.

[56] a) L. Rosgaard, S. Pedersen, J. Langston, D. Akerhielm, J. R. Cherry, A. S. Meyer, *Biotechnol. Prog.* **2007**, *23*, 1270 – 1276; b) A. S. Meyer, L. Rosgaard, H. R. Sørensen, *J. Cereal Sci.* **2009**, *50*, 337 – 344.

[57] C. R. South, D. A. L. Hogsett, L. R. Lynd, *Enzyme Microb. Technol.* **1995**, *17*, 797 – 803.

[58] a) X. Shao, L. Lynd, C. Wyman, *Biotechnol. Bioeng.* **2009**, *102*, 66 – 72; b) X. Shao, L. Lynd, C. Wyman, A. Bakker, *Biotechnol. Bioeng.* **2009**, *102*, 59 – 65.

[59] R. E. T. Drissen, R. H. W. Maas, M. J. E. C. Van Der Maarel, M. A. Kabel, H. A. Schols, J. Tramper, H. H. Beeftink, *Biocatal. Biotransform.* **2007**, *25*, 419 – 429.

[60] a) S. T. Merino, J. Cherry, *Adv. Biochem. Eng./Biotechnol.* **2007**, *108*, 95 – 120; b) S. Brethauer, C. E. Wyman, *Bioresour. Technol.* **2010**, *101*, 4862 – 4874; c) M. Galbe, P. Sassner, A. Wingren, G. Zacchi, *Adv. Biochem. Eng./Biotechnol.* **2007**, *108*, 303 – 327; d) M. Galbe, G. Zacchi, *Adv. Biochem. Eng./Biotechnol.* **2007**, *108*, 41 – 65.

[61] D. Cannella, H. Jørgensen, *Biotechnol. Bioeng.* **2014**, *111*, 59 – 68.

[62] a) M. M. Bomgardner, *Chem. Eng. News* **2013**, *91*(4), 20 – 22; b) M. M. Bomgardner, *Chem. Eng. News* **2012**, *90*(5), 11; c) M. M. Bomgardner, *Chem. Eng. News* **2012**, *90*(24), 18 – 21.

[63] M. McCoy, *Chem. Eng. News* **2012**, *90*(41), 22 – 25.

[64] *Nature* **2010**, *463*, 592.

[65] *Chem. Eng. News* **2012**, *90*(24), 17.

[66] M. M. Bomgardner, *Chem. Eng. News* **2013**, *91*(51), 9.

[67] M. Rarbach, Y. Soltl, *Chimia* **2013**, *67*, 732 – 734.

[68] R. Vanholme, I. Cesarino, K. Rataj, Y. Xiao, L. Sundin, G. Goeminne, H. Kim, J. Cross, K. Morreel, P. Araujo, L. Welsh, J. Haustraete, C. McClellan, B. Vanholme, J. Ralph, G. G. Simpson, C. Halpin, W. Boerjan, *Science* **2013**, *341*, 1103 – 1106.

[69] D. B. Wilson, *Curr. Opin. Microbiol.* **2011**, *14*, 259 – 263.

[70] a) R. I. Amann, W. Ludwig, K. H. Schleifer, *Microbiol. Rev.* **1995**, *59*, 143 – 169; b) P. Hugenholtz, B. M. Goebel, N. R. Pace, *J. Bacteriol.* **1998**, *180*, 4765 – 4774; c) N. R. Pace, *Science* **1997**, *276*, 734 – 740; d) J. T. Staley, A. Konopka, *Annu. Rev. Microbiol.* **1985**, *39*, 321 – 346.

[71] a) W. R. Streit, R. Daniel, *Metagenomics*, Springer Protocols, Berlin, **2010**; b) P. Lorenz, J. Eck, *Nat. Rev. Microbiol.* **2005**, *3*, 510 – 516.

[72] a) L. Fernández-Arrojo, M.-E. Guazzaroni, N. López-Cortés, A. Beloqui, M. Ferrer, *Curr. Opin. Biotechnol.* **2010**, *21*, 725 – 733; b) T. Uchiyama, K. Miyazaki, *Curr. Opin. Biotechnol.* **2009**, *20*, 616 – 622; c) M. Ferrer, A. Beloqui, K. N. Timmis, P. N. Golyshin, *J. Mol. Microbiol. Biotechnol.* **2009**, *16*, 109 – 123; d) L. L. Li, S. R. McCorkle, S. Monchy, S. Taghavi, D. van der Lelie, *Biotechnol. Biofuels* **2009**, *2*, 10; e) C. Simon, R. Daniel, *Appl. Microbiol. Biotechnol.* **2009**, *85*, 265 – 276; f) I. André, G. Potocki-Véronèse, S. Barbe, C. Moulis, M. Remaud-Siméon, *Curr. Opin. Chem. Biol.* **2014**, *19*, 17 – 24.

[73] L. Tasse, J. Bercovici, S. Pizzut-Serin, P. Robe, J. Tap, C. Klopp, B. L. Cantarel, P. M. Coutinho, B. Henrissat, M. Leclerc, J. Doré, P. Monsan, M. Remaud-Simeon, G. Potocki-Veronese, *Genome Res.* **2010**, *20*, 1605 – 1612.

[74] a) H. J. Flint, E. A. Bayer, M. T. Rincon, R. Lamed, B. A. White, *Nat. Rev. Microbiol.* **2008**, *6*, 121 – 131; b) C. A. Lozupone, M. Hamady, B. L. Cantarel, P. M. Coutinho, B. Henrissat, J. I. Gordon, R. Knight, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15076 – 15081.

[75] D. A. Cecchini, E. Laville, S. Laguerre, P. Robe, M. Leclerc, J. Doré, B. Henrissat, M. Remaud-Siméon, P. Monsan, G. Potocki-Véronèse, *PLoS One* **2013**, *8*, e72766.

[76] R. Irague, S. Massou, C. Moulis, O. Saurel, A. Milon, P. Monsan, M. Remaud-Siméon, J. C. Portais, G. Potocki-Véronèse, *Anal. Chem.* **2011**, *83*, 1202 – 1206.
[77] M. Hess, A. Sczyrba, R. Egan, T. W. Kim, H. Chokhawala, G. Schroth, S. Luo, D. S. Clark, F. Chen, T. Zhang, R. I. Mackie, L. A. Pennacchio, S. G. Tringe, A. Visel, T. Woyke, Z. Wang, E. M. Rubin, *Science* **2011**, *331*, 463 – 467.
[78] S. Wongwilaiwalin, T. Laothanachareon, W. Mhuantong, S. Tangphatsornruang, L. Eurwilaichitr, Y. Igarashi, V. Champreda, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 8941 – 8954.
[79] a) S. E. Blumer-Schuette, I. Kataeva, J. Westpheling, M. W. Adams, R. M. Kelly, *Curr. Opin. Biotechnol.* **2008**, *19*, 210 – 217; b) A. Labourel, M. Jam, A. Jeudy, J. H. Hehemann, M. Czjzek, G. Michel, *J. Biol. Chem.* **2014**, *289*, 2027 – 2042.
[80] J. E. Graham, M. E. Clark, D. C. Nadler, S. Huffer, H. A. Chokhawala, S. E. Rowland, H. W. Blanch, D. S. Clark, F. T. Robb, *Nat. Commun.* **2011**, *2*, 375.
[81] a) V. G. Eijsink, G. Vaaje-Kolstad, K. M. Varum, S. J. Horn, *Trends Biotechnol.* **2008**, *26*, 228 – 235; b) S. J. Horn, G. Vaaje-Kolstad, B. Westereng, V. G. Eijsink, *Biotechnol. Biofuels* **2012**, *5*, 45.
[82] a) K. Neufeld, J. Pietruszka, *ChemCatChem* **2012**, *4*, 1239 – 1240; b) S. Fushinobu, *Nat. Chem. Biol.* **2014**, *10*, 88 – 89.
[83] R. J. Quinlan, M. D. Sweeney, L. Lo Leggio, H. Otten, J. C. Poulsen, K. S. Johansen, K. B. Krogh, C. I. Jorgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C. P. Walter, P. Dupree, F. Xu, G. J. Davies, P. H. Walton, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15079 – 15084.
[84] a) C. M. Phillips, W. T. Beeson, J. H. Cate, M. A. Marletta, *ACS Chem. Biol.* **2011**, *6*, 1399 – 1406; b) W. T. Beeson, C. M. Phillips, J. H. Cate, M. A. Marletta, *J. Am. Chem. Soc.* **2012**, *134*, 890 – 892; c) X. Li, W. T. Beeson, C. M. Phillips, M. A. Marletta, J. H. Cate, *Structure* **2012**, *20*, 1051 – 1061.
[85] G. R. Hemsworth, B. Henrissat, G. J. Davies, P. H. Walton, *Nat. Chem. Biol.* **2014**, *10*, 122 – 126.
[86] G. Vaaje-Kolstad, B. Westereng, S. J. Horn, Z. Liu, H. Zhai, M. Sorlie, V. G. Eijsink, *Science* **2010**, *330*, 219 – 222.
[87] L. M. Chauvigné-Hines, L. N. Anderson, H. M. Weaver, J. N. Brown, P. K. Koech, C. D. Nicora, B. A. Hofstad, R. D. Smith, M. J. Wilkins, S. J. Callister, A. T. Wright, *J. Am. Chem. Soc.* **2012**, *134*, 20521 – 20532.
[88] a) T. Davids, M. Schmidt, D. Böttcher, U. T. Bornscheuer, *Curr. Opin. Chem. Biol.* **2013**, *17*, 215 – 220; b) U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185 – 194; c) N. J. Turner, *Nat. Chem. Biol.* **2009**, *5*, 567 – 573; d) R. J. Kazlauskas, U. T. Bornscheuer, *Nat. Chem. Biol.* **2009**, *5*, 526 – 529; e) S. Lutz, U. T. Bornscheuer, *Protein Engineering Handbook*, Vol. 1 & 2, Wiley-VCH, Weinheim, **2009**; f) S. Lutz, U. T. Bornscheuer, *Protein Engineering Handbook*, Vol. 3, Wiley-VCH, Weinheim, **2012**; g) F. H. Arnold, G. Georgiou in *Methods Mol. Biol.*, *Vol. 231*, Humana, Totawa, **2003**; h) M. W. Peters, P. Meinhold, A. Glieder, F. H. Arnold, *J. Am. Chem. Soc.* **2003**, *125*, 13442 – 13450.
[89] R. C. Cadwell, G. F. Joyce, *PCR Methods Appl.* **1992**, *2*, 28 – 33.
[90] W. P. C. Stemmer, *Nature* **1994**, *370*, 389 – 391.
[91] M. Schülein, *Biochim. Biophys. Acta Struct. Mol. Enzymol.* **2000**, *1543*, 239 – 252.
[92] A. Rentmeister, *Chem. Ing. Tech.* **2013**, *85*, 818 – 825.
[93] W. J. Liu, X. Z. Zhang, Z. M. Zhang, Y. H. P. Zhang, *Appl. Environ. Microbiol.* **2010**, *76*, 4914 – 4917.
[94] M. M. Meyer, L. Hochrein, F. H. Arnold, *Prot. Eng. Des. Sel.* **2006**, *19*, 563 – 570.
[95] P. C. Bruijnincx, B. M. Weckhuysen, *Angew. Chem.* **2013**, *125*, 12198 – 12206; *Angew. Chem. Int. Ed.* **2013**, *52*, 11980 – 11987.
[96] L. Lin, X. Meng, P. F. Liu, Y. Z. Hong, G. B. Wu, X. L. Huang, C. C. Li, J. L. Dong, L. Xiao, Z. Liu, *Appl. Microbiol. Biotechnol.* **2009**, *82*, 671 – 679.
[97] V. H. Vu, K. Kim, *J. Microbiol. Biotechnol.* **2012**, *22*, 607 – 613.
[98] M. Anbar, O. Gul, R. Lamed, U. O. Sezerman, E. A. Bayer, *Appl. Environ. Microbiol.* **2012**, *78*, 3458 – 3464.
[99] T. Fukuda, T. Ishikawa, M. Ogawa, S. Shiraga, M. Kato, S. I. Suye, M. Ueda, *Biotechnol. Prog.* **2006**, *22*, 933 – 938.
[100] Y. S. Kim, H. C. Jung, J. G. Pan, *Appl. Environ. Microbiol.* **2000**, *66*, 788 – 793.
[101] a) P. Heinzelman, C. D. Snow, M. A. Smith, X. L. Yu, A. Kannan, K. Boulware, A. Villalobos, S. Govindarajan, J. Minshull, F. H. Arnold, *J. Biol. Chem.* **2009**, *284*, 26229 – 26233; b) P. Heinzelman, C. D. Snow, I. Wu, C. Nguyen, A. Villalobos, S. Govindarajan, J. Minshull, F. H. Arnold, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5610 – 5615.
[102] I. Wu, F. H. Arnold, *Biotechnol. Bioeng.* **2013**, *110*, 1874 – 1883.
[103] Y. Ito, A. Ikeuchi, C. Imamura, *Prot. Eng. Des. Sel.* **2013**, *26*, 73 – 79.
[104] S. P. Voutilainen, H. Boer, M. Alapuranen, J. Janis, J. Vehmaanpera, A. Koivula, *Appl. Microbiol. Biotechnol.* **2009**, *83*, 261 – 272.
[105] S. P. Voutilainen, P. G. Murray, M. G. Tuohy, A. Koivula, *Prot. Eng. Des. Sel.* **2010**, *23*, 69 – 79.
[106] P. Heinzelman, R. Komor, A. Kanaan, P. Romero, X. L. Yu, S. Mohler, C. Snow, F. Arnold, *Prot. Eng. Des. Sel.* **2010**, *23*, 871 – 880.
[107] R. S. Komor, P. A. Romero, C. B. Xie, F. H. Arnold, *Prot. Eng. Des. Sel.* **2012**, *25*, 827 – 833.
[108] C. M. Dana, P. Saija, S. M. Kal, M. B. Bryan, H. W. Blanch, D. S. Clark, *Biotechnol. Bioeng.* **2012**, *109*, 2710 – 2719.
[109] C. Dumon, A. Varvak, M. A. Wall, J. E. Flint, R. J. Lewis, J. H. Lakey, C. Morland, P. Luginbühl, S. Healey, T. Todaro, G. DeSantis, M. Sun, L. Parra-Gessert, X. Tan, D. P. Weiner, H. J. Gilbert, *J. Biol. Chem.* **2008**, *283*, 22557 – 22564.
[110] M. Timm, J. Görl, M. Kraus, S. Kralj, H. Hellmuth, R. Beine, K. Buchholz, L. Dijkhuizen, J. Seibel, *ChemBioChem* **2013**, *14*, 2423 – 2426.
[111] M. A. Smith, A. Rentmeister, C. D. Snow, T. Wu, M. F. Farrow, F. Mingardon, F. H. Arnold, *FEBS J.* **2012**, *279*, 4453 – 4465.