Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Genomics of cellulolytic bacteria

Daniela E Koeck[1,3], Alexander Pechtl[1,3], Vladimir V Zverlov[1,2] and Wolfgang H Schwarz[1]

The heterogeneous plant biomass is efficiently decomposed by the interplay of a great number of different enzymes. The enzyme systems in cellulolytic bacteria have been investigated by sequencing and bioinformatic analysis of genomes from plant biomass degrading microorganisms with valuable insights into the variety of the involved enzymes. This broadened our understanding of the biochemical mechanisms of plant polymer degradation and made the enzymes applicable for modern biotechnology.A list of the truly cellulolytic bacteria described and the available genomic information was examined for proteins with cellulolytic and hemicellulolytic capability. The importance of the isolation, characterization and genomic sequencing of cellulolytic microorganisms and their usage for sustainable energy production from biomass and other residues, is emphasized.

Addresses
[1] Department of Microbiology, Technische Universität München, Emil-Ramann-Str. 4, D-85350 Freising-Weihenstephan, Germany
[2] Institute of Molecular Genetics, Russian Academy of Science, Kurchatov Sq. 2, 123182 Moscow, Russia

Corresponding author: Schwarz, Wolfgang H (wschwarz@wzw.tum.de)

[3] Equally contributing.

Current Opinion in Biotechnology 2014, **29**:171–183

This review comes from a themed issue on **Cell and pathway engineering**

Edited by **Tina Lütke-Eversloh** and **Keith EJ Tyo**

http://dx.doi.org/10.1016/j.copbio.2014.07.002

1367-5931/Published by Elsevier Ltd.

Introduction

The major components of plant biomass, cellulose and hemicellulose, are the most prevalent organic compounds on earth and a tremendous resource of energy [1]. Microbes provide enzymes for the monomerization of the polysaccharides in biomass to sugars which are converted by the same or by other microbes in fermentative processes to the compounds needed.

By contrast to fossil carbon sources, energy produced from biomass is CO_2-neutral. Microbiology can contribute to the CO_2-neutral energy production by supplying efficient polysaccharide hydrolyzing enzymes. In addition ethanol, organic chemicals such as solvents, plastics and core chemicals for generating a huge variety of products can be produced from the sugar derived from plant biomass by microbiological conversion thereby suspending our energy consuming life style.

However, especially cellulose is extremely recalcitrant to enzymatic degradation due to its crystallinity. The ability to hydrolyze it is a rare trait in bacteria and archaea, and is very rarely found in other organisms except in saprophytic fungi. Fungal cellulolytic enzymes occur in the genus *Trichoderma* and *Aspergillus*. They have been extensively investigated and are predominant in applications in the industrial biotechnology [2•]. Recently, there were also some reports about cellulolytic crenarchaeal species, such as *Desulfurococcus fermentans* and *Thermogladius cellulolyticus* [3,4]. However, the knowledge about cellulolytic species within the archaeal kingdom is still very limited.

This review is focusing on the truly cellulolytic bacteria, which are able to utilize natural (=crystalline) cellulose. Besides the cellulolytic fungi these few truly cellulolytic bacteria are the efficient hydrolyzers of plant cell wall polysaccharides in nature, especially in thermophilic anaerobic ecosystems. These saprophytic bacteria are remarkably well adapted and secrete a plethora of enzymes with a distinct synergistic behavior. Their highly specialized, often intricate enzyme systems include cellulases, hemicellulases, pectinases and other related glycoside hydrolases, as well as polysaccharide lyases and carbohydrate esterases [5••]. They expose cellulose fibers and degrade them to a mixture of cellobiose, cellotetraose and other cellodextrins, which are taken up by the cells and further degraded by β-glucosidases/cellobiases, or cellobiose-phosphorylases and cellodextrin-phosphorylases. Concomitantly the hemicelluloses and pectins are degraded as well.

Cellulose degrading bacteria

About half of the bacteria containing genes for cellulases, hemicellulases and pectinases are saprophytes, the bacteria efficiently degrading dead plant biomass (data from March 2011) [6••]. However, only a small number possess more than 3 genes for β-1,4-glucanases (cellulases), a prerequisite for the effective degradation of natural cellulose. In fact, up to date only relatively few bacterial species able to hydrolyze and to utilize natural (=crystalline) cellulose was isolated and characterized [7]. A list of the bacterial species known to utilize cellulose as sole

carbon and energy source is compiled in Table 1, which uses the actual phylogenetic classification (data as of May 2014) [8].

Most bacteria called 'cellulolytic' in the literature produce extracellular β-glucanases, usually endo-glucanases, which cleave the β-1,4-glucosidic bond only in soluble (mixed-linkage) β-glucans or in artificial cellulosic compounds such as carboxymethylcellulose (CMC). Although CMC is chemically speaking cellulose with its typical β-1,4-glucosidic linkages, its depolymerization is only a precondition but not sufficient for degradation of natural (i.e. crystalline) cellulose. Unlike the truly cellulolytic species, these so-called 'cellulolytic' bacteria are unable to totally degrade and utilize crystalline cellulose. Their cellulases are functional, for example, for the infection of plant cells (in plant pathogens), for cellulose synthesis (in bacteria as well as in plants), and for other purposes than the metabolization of crystalline cellulose [6••].

Truly cellulolytic microorganisms are of great importance for the production of sustainable energy: they degrade plant material; they produce a variety of desired chemicals through fermentation of biomass which would be otherwise difficult to use appropriately; and they provide the genetic information for the production of a multitude of cellulolytic and hemicellulolytic enzymes that can be used for application in many industrial and biotechnological processes. Recently the genomic sequences of a number of truly cellulolytic species became publicly available (Table 1; right column) and can now be mined for the presence of potentially useful enzymes and further analyzed for a better understanding of microbial mechanisms to metabolize (hemi-)cellulose [9].

Enzymes degrading polysaccharides in plant biomass

In plant cell walls the most abundant sugar is the glucose in cellulose. Cellulose is a structurally robust polymer because it is organized in large crystals containing tens of thousands of glucose molecules. It is therefore not easily accessible for the attacking enzymes. In addition the cellulose crystals are embedded in (but not covalently linked to) a matrix of hemicellulose and pectin that are themselves rich sources of carbohydrates for saprophytic bacteria through the enzymes they secrete. Hemicellulose has to be removed to increase the cellulose accessibility. Due to the various structures (cellulose) and divers chemical composition (hemicelluloses, pectins), a great number of enzymes with differing substrate specificity are necessary to degrade the polysaccharides in plants [10]. The sequencing of bacterial genomes has revealed the multitude of enzymes involved in the degradation.

The involved enzymes are presented in the CAZy database [11••] in functional categories such as glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities and associated modules (such as carbohydrate binding modules, CBMs) [12]. The GHs contain enzymes that hydrolyze the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate component [12]. The CAZy database classifies them based on the amino acid sequence and the resulting fold of the protein into thus far 133 families [11••,13]. The GH family membership of an enzyme thus provides insight into the comparative structural features and evolutionary relationships with other family members and its enzymatic mechanism [5••].

Cellulases are enzymes degrading β-1,4-glucosidic linkages by a hydrolytic mechanism which is either inverting or retaining the configuration of the anomeric carbon. Enzymes belonging to 6 (8 including cellobiose-degrading β-glucosidases) differently folded glycosyl hydrolase families are known to hydrolyze cellulose. Besides the hydrolytic mechanism, the enzymes can act in at least three different modes of activity: enzymes with endo-mode cut long β-1,4-glucan molecules randomly, enzymes with a processive mode recognize either the reducing or the non-reducing end of the cellulose molecule, and processive endo-glucanases act processively from the newly created ends within a cellulose molecule producing cellobiose or cellotetraose units [14]. New ends are produced for instance in amorphous regions which are more readily accessible. The different types of cellulases work synergistically to completely hydrolyze the crystalline cellulose microfibrils [15•]. Synergism between different enzymes is occurring when the ratio between the combined activity and the sum of the single enzyme activities exceeds 1.0. Most GH-families contain non-processive endo-glucanases as well as processive cellobiohydrolases or processive endo-glucanases [1]. The topology of the active site differs between endo-glucanases and exo-glucanases. Whereas endo-glucanases typically possess a cleft-like topology that allows the access to long molecules and cleavage at a random position within the cellulose chain, the active site of the exo-glucanases has a tunnel-like shape formed by long loops of the protein molecule that fold over the active site residues and cleave cellodextrines of defined size from one of the ends of the substrate chain [16] (Figure 1).

Cellulases and hemicellulases usually have a modular structure, combining a catalytic module with another catalytic module (acting synergistically) and/or with non-catalytic modules such as one or more substrate binding modules [1]. The latter are called carbohydrate binding modules (CBMs) and may have affinity for crystalline or amorphic cellulose, or bind to other polymeric carbohydrates. They influence cellulase activity through attachment to the substrate. Targets may be single, long or short, cellulose molecules, or crystalline structures of different type [17]. By fixing the enzyme to the substrate