ADVANCES IN BIOCHEMICAL ENGINEERING BIOTECHNOLOGY

65

Managing Editor:
T. Scheper

Volume Editor:
G. T. Tsao

Springer

Recent Progress in Bioconversion of Lignocellulosics

Recent Progress in Bioconversion of Lignocellulosics

Volume Editor: G. T. Tsao

With contributions by
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, G. T. Tsao,
J. S. Tolan, R. W. Torget, D. B. Wilson, L. Xia

Springer

Advances in Biochemical Engineering/Biotechnology reviews actual trends in modern biotechnology. Its aim is to cover all aspects of this interdisciplinary technology where knowledge, methods and expertise are required for chemistry, biochemistry, microbiology, genetics, chemical engineering and computer science. Special volumes are dedicated to selected topics in which the interdisciplinary interactions of this technology are reflected. New biotechnological products and new processes for synthesizing and purifying these products are at the center of interest. New discoveries and applications are discussed.

In general, special volumes are edited by well known guest editors. The managing editor and publisher will however always be pleased to receive suggestions and supplementary information. Manuscripts are accepted in English.

In references Advances in Biochemical Engineering/Biotechnology is abbreviated as Adv. Biochem. Engin./Biotechnol. as a journal.

ISSN 0724-6145
ISBN 3-540-65577-8
Springer-Verlag Berlin Heidelberg New York

Library of Congress Catalog Card Number 72-152360

Typesetting: Fotosatz-Service Köhler GmbH, Würzburg
Cover: Design & Production, Heidelberg
SPIN: 10648397 02/3020 – 5 4 3 2 1 0 – Printed on acid-free paper

Managing Editor

Professor Dr. T. Scheper
Institute of Technical Chemistry
University of Hannover
Callinstraße 3
D-30167 Hannover/FRG
E-mail: scheper@mbox.iftc.uni-hannover.de

Volume Editor

Prof. Dr. G. T. Tsao
Laboratory of Renewable
Resources Engineering
1295 Potter Center, Room 216
Purdue University
West Lafayette, IN 47907-1295/USA
E-mail: tsaogt@ecn.purdue.edu

Editorial Board

Springer

Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Attention all "Enzyme Handbook" Users:

Information on this handbook can be found via the internet at

http://www.springer.de/chem/samsup/enzym-hb/ehb_home.html

At no charge you can download the complete volume indexes Volumes 1 through 17 from the Springer www server at the above mentioned URL. Just click on the volume you are interested in and receive the list of enzymes according to their EC-numbers.

Preface

This volume describes recent advances in the bioconversion of lignocellulosics. It starts with two articles on genetics and properties of cellulases and their reaction kinetics and mechanisms. The cost of cellulases has been a hindrance to large scale use of enzymatic hydrolysis. Two articles on cellulase production by submerged fermentation and by solid state fementation are included to describe the state of the art in this area. Dilute acid hydrolysis of cellulose continues to be of interest as well as potentially useful. The most recent advances in this area is also covered. A great deal of progress has been made in genetic engineering for improved regulation of xylose fermentation by yeasts. An article on genetically engineered *Saccharomyces* for simulteaneous fermentation of glucose and xylose describes the importance advances made in production of fuel ethanol from lignocellulosic biomass. In recent years, there has been increasing interests in recycling and the reuse of scrap paper as well as environment considerations. A contribution is presented which describes the research perspectives in that area. Finally, recent advances in the use of lignocellulosic biomass for the production of ethanol and organic acids are presented in two articles.

Renewable resources are inevitably of great importance in the years to come. There is a never-ending search for better living conditions for human beings. The more resource materials can be recycled, the richer we will be. Bioconversion of lignocellulosics, natural and man-made, is an important link in that cycle. Extensive use of renewable resources will also slow down continued deterioration of the environment.

Advances are being made as this volume is being put together. Another volume on the same subject, perhaps, should be prepared in another ten years or ever sooner.

March 1999 George T. Tsao

65

Advances in Biochemical Engineering/Biotechnology

Managing Editor: T. Scheper

Contents

Genetics and Properties of Cellulases

David B. Wilson · Diana C. Irwin

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853, USA, e-mail: dbw3@cornell.edu

Cellulases are enzymes which degrade the insoluble, abundant polymer cellulose. In order to perform this task bacteria, fungi, plants and insects have developed a variety of different systems with multiple cellulases. In this review the similarities and differences of these enzymes are summarized based on the burgeoning information gained in recent years from amino acid sequences, three dimensional structures and biochemical experiments. The independent cellulases of aerobic organisms are contrasted with the cellulosomes of anaerobic organisms. The ability of different enzymes to synergize with each other is discussed along with the role of the different types of enzymes in cellulose degradation.

Keywords. Cellulosome, Endoglucanase, Cellobiohydrolase, Synergism, Mechanism, Regulation, Structure, Application

Advances in Biochemical Engineering/ Biotechnology, Vol. 65
Managing Editor: Th. Scheper

List of Abbreviations

CBD cellulose-binding domain
CMC carboxymethyl cellulose
SC phosphoric acid swollen cellulose

1 Introduction

Cellulose is the most abundant polymer on earth with an estimated 10^{12} metric tons produced each year by plants [1] and cellulases produced by fungi and bacteria are responsible for most cellulose degradation [2]. Cellulose is a linear homopolymer of β1–4 linked glucose residues. There are stereochemical differences between adjacent glucose residues so that the repeating unit in cellulose is the disaccharide, cellobiose, and the main product of the enzymatic hydrolysis of cellulose is cellobiose. Cellulose is difficult to degrade because cellulose molecules can form tightly packed, extensively hydrogen-bonded regions called crystalline cellulose [3, 4]. The crystalline regions are believed to be separated by less ordered amorphous regions, but these still contain many hydrogen bonds. Cellulose is insoluble and oligomers containing more than six residues are also insoluble. The resistance of cellulose to degradation may be responsible for the large number and types of cellulases produced by cellulose degrading organisms. There are two basic types of cellulases, endo- and exocellulases. Endocellulases have a more open active site cleft and can bind at any available point along a cellulose molecule hydrolyzing a few bonds before dissociation. In contrast, exocellulases, also called cellobiohydrolases, have an active site tunnel and can only access the ends of a cellulose molecule [5] cleaving off cellobiose processively. There are several informative and detailed reviews on various aspects of cellulases [2, 6–9]. This paper is an overview of the remarkable variety of cellulases and discusses the similarities and differences in their properties.

1.1 Eukaryotic and Prokaryotic Cellulases

Eukaryotic cellulases have been found in insects, plants and fungi while bacteria producing prokaryotic cellulases are found wherever cellulose is present such as in compost piles, soil, rotting wood, etc. Many animals, including ruminants, utilize the cellulose present in their food; however, they do not produce cellulases but rely on cellulolytic microorganisms, primarily bacteria, to hydrolyze the cellulose [10]. The rumen is an extremely anaerobic environment so all rumen organisms are strict anaerobes. Some insects, for example termites and wood roaches, degrade cellulose. At first it was thought that insects did not produce cellulases and that symbiotic cellulolytic bacteria, fungi or protozoa produced the cellulases they used for cellulose degradation. However, there is growing evidence that several different species of insects produce cellulases and thus contain cellulase genes. Insect cellulases have been isolated from four species of termites, *Macrotermes subhyalinus*, *M. michaelseni*, *Eoptotermes lactens* and

Natsutitermes walheri, as well as from the wood roach, *Panesthia cribrata* [11]. A gene encoding a family 9 endocellulase has been cloned from the termite *Reticulitermes speratus*, which contains a 450 base intron, proving it is a eukaryotic gene [12].

Cellulases are also produced by plants and participate in leaf and flower abscission, the ripening of fruits, as well as differentiation of vascular tissue and plant cell wall growth [13]. A number of plant cellulase genes have been cloned and sequenced. All of them belong to cellulase family 9 and none of them have been shown to contain a cellulose-binding domain (CBD) [14, 15]. Both of these results are surprising since the presence of a CBD appears to be important for degrading crystalline cellulose and most cellulolytic organisms contain cellulase genes from several cellulase families. The methods used to clone plant cellulases should have detected enzymes from any family. Most plants contain multiple cellulase genes and they appear to be regulated in different ways. The cellulases that are involved in fruit ripening are often induced by ethylene, while the cellulases involved in cell growth are often induced by auxin [16, 17].

In addition, a new class of proteins (α-expansins) has been discovered in plants that appear to disrupt the interactions between cellulose chains, without hydrolytic activity, to allow cell wall expansion [18]. A different, but related, set of molecules (β-expansins) has been found in pollen [19]. These molecules appear to help the growing pollen tube penetrate through the cell walls in the ovule allowing fertilization. There is preliminary evidence that the addition of expansin to cellulases stimulates crystalline cellulose hydrolysis [20].

Endocellulases have been isolated from plant pathogenic nematodes and four structural genes were cloned from two species [21]. All of the cellulase catalytic domains belong to family 5 and two of them also code for family II CBDs. The enzymes show 37% identity in their amino acid sequence to several bacterial cellulases.

1.2 Anaerobic Versus Aerobic Cellulases

There are two quite different ways that cellulolytic bacteria and fungi deal with the problem raised by the insolubility of cellulose and their inability to ingest cellulose particles. Most anaerobic microorganisms produce multienzyme complexes, called cellulosomes, on their cell surface while most aerobic microorganisms secrete a set of individual cellulases into the external milieu where the enzymes act synergistically to degrade crystalline cellulose. In each case, the products of digestion are oligosaccharides, mostly cellobiose and glucose, that are transported into the cell and metabolized. Some aerobic fungi also secrete cellobiase so that glucose is the major end product of cellulose degradation.

One possible explanation for the difference in cellulase organization is that anaerobic organisms are more energy limited than aerobic organisms, and thus it is more important for them to retain the products of cellulose digestion. Some anaerobic organisms are tightly bound to cellulose by their surface cellulases so that the products of digestion are released in a confined space between the insoluble cellulose and the organism. With free cellulases, the hydrolysis products

are in solution and would be more available to competing organisms. Since there is an anaerobic bacterium, *Clostridium papyrosolvens* C7, which secretes cellulase complexes into the medium and thus would not be able to retain all of the digestion products, there may be other advantages of complex formation [22, 23].

2
Cellulase Domains

Cellulases usually have several domains. All of them contain one catalytic domain and a few with multi-catalytic domains have been found. A lambda recombinant from a genomic library of *Caldocellum saccharolyticum* encoded three multi-catalytic domain enzymes: CelA with family 9 and family 48 cellulase catalytic domains, ManA with β-mannanase and endocellulase catalytic domains, and CelB with xylanase and endocellulase catalytic domains [24–26]. *Anaerocellum thermophilum* CelA has both a family 9 and a family 48 catalytic domain [27]. *Clostridium thermocellum* CelJ has a family 5 endocellulase domain combined with a family 9 processive endocellulase domain [28]. There have not yet been detailed studies into whether the activity of a cellulase with two catalytic domains is higher than the activity of a mixture of enzymes containing the two domains by themselves. There is a report that an enzyme containing a family 5 domain and a *Bacillus* 1–3, β1–4 glucanase domain has higher activity on β-glucan than either of the domains alone [29].

After catalytic domains, the next most common domains are CBDs which are usually joined to the catalytic domain by a short linker peptide. Cellulases that are present in cellulosomes contain short domains called dockerins that bind to specific sites on a scaffoldin protein to form a cellulosome [30]. A number of cellulases contain fibronectin-like domains, but the function of these domains is not known [6]. There are several other domains with unknown functions [6].

2.1
Catalytic Domain Families

Many cellulase genes have been cloned and sequenced. At this time, there are at least 112 sequences reported in the Swiss protein data base. Henrissat and colleagues have grouped the bulk of these genes into eleven families (5–9, 12, 44, 45, 48, 60, 61) based on both sequence homology and hydrophobic cluster analysis [31–34]. There is a web site at http://expasy.hcuge.ch/cgi-bin/lists?glycosid.txt which contains current information on all glycosyl hydrolase families including cellulases. In addition, several cellulase genes code for enzymes that do not resemble any other known cellulases. The presence of a large number of cellulase families is unusual as most enzymes have only a few families. This heterogeneity presumably results from the abundance of cellulose, the complexity and variability of plant cell walls, which are the actual substrates of most cellulases, and the difficulty of degrading plant cell wall cellulose.

X-ray structures have been determined for more than a dozen cellulases from eight different families. The results support the idea that all of the cellulases in a given family have the same basic structure [35–49]. Gideon Davies has

written a concise overview of the basic structures [50]. At least six completely different folds can lead to an active cellulase. Some cellulase families include both exocellulase and endocellulase genes, while others contain only one type of enzyme. Despite the dramatic difference in the way their cellulases are organized (free versus bound), the families to which the catalytic domains of aerobic and anaerobic cellulases belong show a great deal of overlap. Furthermore, there is some overlap between the families to which bacterial and fungal cellulases belong, although there are families with cellulases from only one class of organism. The one property that is completely conserved in all members of a family is the stereochemistry of cleavage (retaining or inverting) of the cellulose β1-4 bond (Fig. 1) [2, 51, 52] which is discussed in Sect. 5.

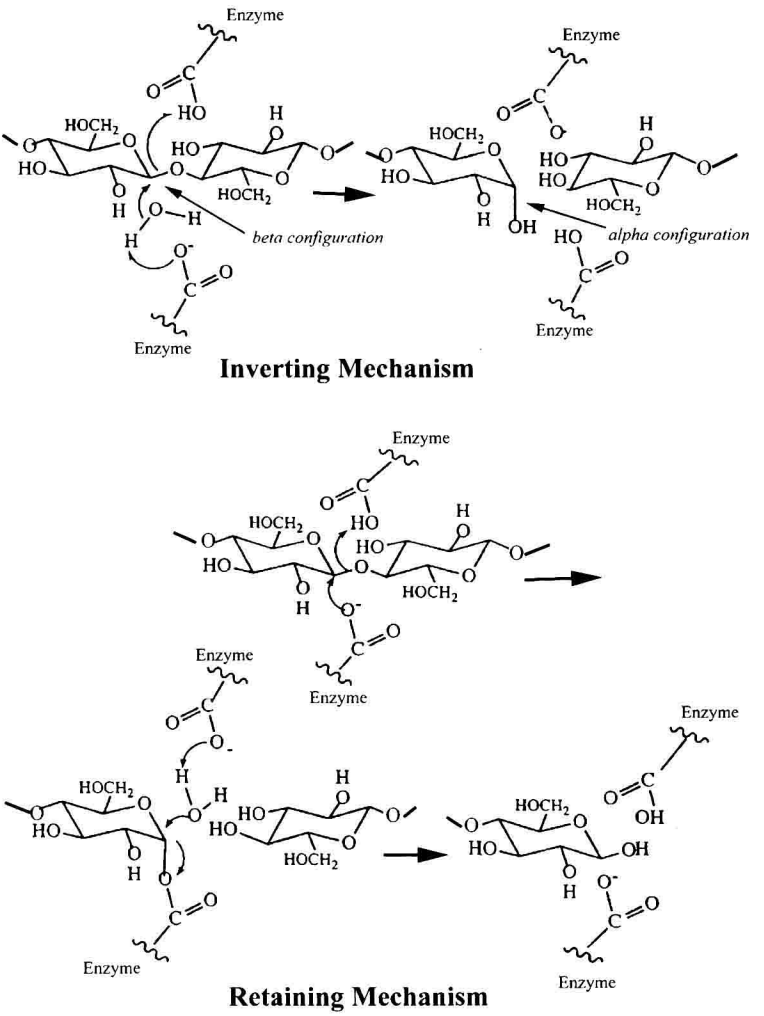

Fig. 1. The two stereochemically different mechanisms of hydrolysis for cellulases

Family 5 is the largest cellulase family containing 57 genes. Of these 52 code for endocellulases and they are retaining enzymes. The other five genes code for β1–3 exoglucanases. Most of the family 5 genes are from bacteria, but some are from fungi. The three-dimensional structures of four family 5 catalytic domains have been reported and they have an α/β-barrel fold, which is the most common fold found among all proteins [40, 44, 53, 54]. Many of the family 5 genes do not code for a cellulose-binding domain. Those that also lack a dockerin domain may not function in the degradation of crystalline cellulose. This has been shown to be true for the carboxymethylcellulase (CMCase) from the anaerobic bacterium *Prevotella bryantii*, where the gene appears to be required for growth on β-glucan, a glucose polymer with alternating β1–4,1–3 linkages [55].

Family 6 contains nine genes coding for both endo- and exocellulases from bacteria and fungi. The structures of two family 6 enzymes, *Trichoderma reesei* CBHII – a fungal exocellulase and *Thermomonospora fusca* E2 – an actinomycete endocellulase, have been determined and they are modified α/β barrels that close the barrel in a slightly different way than it is closed in standard α/β-barrel proteins [43, 46]. When the three-dimensional structures of CBHII and E2 are overlaid there are two loops which cover the active site cleft to form an active site tunnel in CBHII [43]. The E2 active site cleft is much more open because one loop is much shorter and the other has a different conformation. The enzymes in this family all catalyze hydrolysis with inversion of the anomeric carbon configuration.

Family 7 contains only fungal genes, coding for both endo- and exocellulases. The enzymes in this family all utilize the retaining mechanism. The structure of the *T. reesei* CBHI catalytic domain has been determined and it is a unique structure with a β sandwich forming the active site and many loops connecting the β strands [38, 39]. This structure is larger than those of the family 5 and 6 catalytic domains and has a long active site tunnel with enough room for seven glucosyl residues. Three more family 7 structures have been solved: *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable thiooligosaccharide substrate analogue [56], *Humicola insolens* endoglucanase I [57], and *T. reesei* endoglucanase I [58].

Family 8 contains nine genes, all bacterial, which appear to code for endocellulases utilizing the inverting mechanism. The structure of a family 8 endocellulase, *C. thermocellum* Cel A, has been determined and it is an $(\alpha/\alpha)_6$ barrel similar to those found in families 9 and 48 [35, 59].

Family 9 contains 19 cellulase genes that belong to two subfamilies distinguished by the presence or absence of a family III CBD closely attached to the catalytic domain. All of the enzymes in this family are inverting. The three-dimensional structures of the catalytic domain of *C. thermocellum* CelD [41] and the catalytic domain plus the family III CBD of *T. fusca* E_4 [45] have been determined and the catalytic domains are $(\alpha/\alpha)_6$ barrels. The E4 family III CBD was aligned with the catalytic cleft so that a cellulose molecule bound in the active site could also be bound to the CBD. The enzymes without the attached CBD are all endocellulases, while E4, with the attached CBD, is a processive endoglucanase [41, 60]. This family does not contain any fungal genes, but includes genes from both bacteria and plants.

Family 12 contains nine genes, all coding for retaining endoglucanases. They include bacterial and fungal genes. The structure of an endocellulase has been determined and it is a jelly roll made up of β-sheets very similar to the structure of a family 11 xylanase [47].

Families 44, 60 and 61 are small cellulase containing families with only a few members thus far. Family 44 includes an inverting endoglucanase and a mannanase from bacteria. No structures have been determined for these families.

Family 45 contains five endocellulases from bacteria and fungi. The structure of *Humicola insolens* endoglucanase V has been solved and consists of a six-stranded β-barrel domain with a long open groove across the surface. This enzyme catalyzes hydrolysis with inversion at the anomeric carbon atom [36, 37, 61].

Family 48 contains six cellulase genes and they all code for inverting enzymes. Some of the enzymes studied so far appear to have low specific activities on cellulose substrates and several are present in their respective organisms in relatively large amounts suggesting they are exocellulases [25, 27, 62–66]. These genes are present in both anaerobic and aerobic bacteria as well as anaerobic fungi. The three-dimensional structure of one family 48 catalytic domain, *C. cellulolyticum* CelF, has been reported and it is an $(\alpha/\alpha)_6$ barrel similar to that found for family 8 and family 9 cellulases [67]. As expected for an exocellulase part of the active site is in a tunnel.

2.2 Cellulose-Binding Domains

Most anaerobic cellulases either have no cellulose-binding domain (CBD) or, as in *C. thermocellum* cellulosomes, the CBD is attached to the scaffoldin molecule which is in turn attached to multiple catalytic domains. Many aerobic organisms make cellulases with a CBD attached to the catalytic domain via a flexible linker which is often glycosylated. The CBD is usually found at either the N- or the C-terminus in nearly equal numbers. So far, 13 different cellulose-binding domain families have been reported based on sequence differences [6, 68, 69]. Family I CBDs are found only in fungi and are 33–36 amino acids long. Family II CBDs are found only in bacteria and are about 100 residues long. Removal of family II CBDs from *T. fusca* cellulases reduces their activity on crystalline cellulose severely, but affects activity on more soluble or amorphous substrates such as carboxymethyl cellulose (CMC) and phosphoric acid swollen cellulose (SC) much less [70–72]. This has also been shown for removal of the fungal family I CBD from *T. reesei* CBHI and CBHII [73]. Some xylanases contain family II CBDs that can bind to xylan as well as cellulose, but most CBDs do not appear to bind xylan [74, 75]. Family IIIa CBDs are known to anchor the *Clostridium* scaffoldin strongly to cellulose [30]. The *T. fusca* E4 family IIIc CBD has been shown to facilitate processivity of the cellulose molecule through the catalytic active site but does not bind tightly to cellulose [60]. The family IV CBD from *C. fimi* CenC has been shown to bind only to SC and not to crystalline cellulose and has a binding cleft rather than a binding face, enabling it to bind to single cellulose molecules [68, 76].