PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

# The Application of Recombinant-DNA Technology to Cellulases and Lignocellulosic Wastes

J. Knowles, P. Lehtovaara, T. Teeri, M. Penttila, I. Salovuori and L. Andre

*Phil. Trans. R. Soc. Lond. A* 1987 **321**, 449-454
doi: 10.1098/rsta.1987.0022

*Phil. Trans. R. Soc. Lond.* A **321**, 449–454 (1987)
*Printed in Great Britain*



# The application of recombinant-DNA technology to cellulases and lignocellulosic wastes

By J. Knowles, P. Lehtovaara, T. Teeri, M. Penttilä, I. Salovuori and L. André

*Biotechnical Laboratory, VTT, SF-02150 Espoo, Finland*

The efficient utilization of different lignocellulosic wastes will require an understanding of the roles of the different enzymes involved in enzymatic hydrolysis. Towards this goal, we have isolated and characterized the genes coding for four major cellulases, CBH I, CBH II, EG I and EG III, produced by the cellulolytic fungus *Trichoderma reesei*. It seems that *T. reesei* produces at least two classes of cellulase and in each class both endo- and exo-type enzyme are found. Full-length cDNAs coding for CBH I, CBH II, EG I and EG III have been expressed in the yeast *Saccharomyces cerevisiae*, and the cellulases are secreted by these strains. The recombinant cellulases produced by the yeast all show activity towards cellulosic substrates. The information obtained from the cloned genes makes possible the construction of new cellulolytic organisms and may also make possible the development of improved enzymes.

## Introduction

Over the last ten years, the development of a range of different techniques for studying and manipulating the genetic material of living organisms has had a major impact on the biological sciences. Perhaps one of the most important discoveries to come from the application of recombinant-DNA technology was the realization of how little was known about the way in which living systems function. There is now a flood of new information that forces a continuous review of our somewhat simplistic models of the way in which life works.

The ability to isolate, characterize and modify genes has become a very powerful tool for the study of living organisms and their functional elements. The major limitation of these techniques is that because they concern directly the properties of macromolecules, such as the gene (DNA) and the enzyme (polypeptide), a great deal of information concerning the molecular biology and biochemistry of a given system must be available before recombinant-DNA techniques can be fruitfully applied.

Lignocellulose, either in its native forms, or in any of the variety of partly processed forms, is an extremely complex substrate. Lignocellulose contains three main components, lignin, hemicellulose and cellulose. Each of these is closely associated and sometimes covalently linked to the others to form a variable and very complex material. Because of its importance to the paper industry, the physical structure of wood has been extensively studied (Sjöström 1981), but knowledge concerning the details of the molecular structure of this material is still largely missing. This is (at least in part) due to the non-specific nature of the chemical methods used to investigate the fine structure of lignocellulose. Cellulose is chemically the simplest of the three components, as it is simply a β-1,4-linked glucose polymer, but its degree of crystallinity is highly variable depending on the source (Henrissat *et al.* 1985). In addition to the variation of the

ratios of the three different components even on a microscale in the plant, the two other elements, lignin and hemicellulose, can vary in chemical composition and in physical properties.

Despite the complexity of the lignocellulose 'substrate', it is broken down efficiently in Nature by a multitude of microorganisms, sometimes with help from macroorganisms such as ruminants. Each organism appears to be specialized to play a particular role in this breakdown process and clearly synergy between different organisms is necessary for efficient breakdown. It would also seem that each organism is producing a wide range of degradative enzymes. The understanding of the biodegradation of lignocellulose therefore represents a major challenge because the substrate is extremely complex, the degradative microorganisms are many, and each organism apparently produces many degradative enzymes. However, because lignocellulose is the most important renewable resource on this planet, it is important to accept this challenge. A better understanding of the nature of this resource will clearly lead to new possibilities for the utilization of the multitude of different forms available now and also for those that become available in the future.

## Fungal cellulases

To obtain more information concerning the molecular biology of lignocellulose degradation, we have chosen to study the genes and the enzymes of the efficient cellulolytic system produced by the filamentous fungus *Trichoderma reesei*. There are a number of reasons for this choice. The cellulases from *T. reesei* are able to function individually and some purified enzymes are able to hydrolyse crystalline cellulose alone. Thus they differ from bacterial cellulases, which according to the available evidence degrade crystalline cellulose only when working as a physical complex of many different enzymes.

Despite the fact that much effort has been devoted to the study of the enzymology of cellulases for a number of years, the substrate and product specificity and even mode of action of particular enzymes is still not completely clear (Enari & Niku-Paavola 1987). While not wishing to review in detail the work on the biochemistry of cellulases, we present some of the reasons why it is difficult to study the enzymology of cellulases.

### *The substrate*

Cellulose is chemically simple but physically complex, containing different types of crystal and a varying amount of amorphous structure. Although all covalent bonds are of the β-1,4-glycosidic type, the microenvironment surrounding each bond may vary considerably. It is often very difficult to standardize the substrate and to compare results obtained in different laboratories. The substrate is insoluble and the question of exactly which parameter should be measured during hydrolysis is crucial. An estimate of the amount of material lost from the solid may be different from the amount of sugar solubilized. The relation between the enzymatic hydrolysis of small synthetic substrates such as 4-methylumbelliferyl glycosides and hydrolysis of crystalline cellulose is not yet clear (van Tilbeurgh *et al.* 1982). One of the major problems is that the nature of the substrate changes during hydrolysis. Initially, amorphous non-crystalline regions are attacked, because they are more easily hydrolysed. Later, as the percentage of crystalline cellulose in the sample rises, the rate of hydrolysis falls. The rate at which these changes occur is very difficult to control and depends very much on which enzymes are being studied.





PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

### *The enzymes*

Fungal cellulolytic enzymes have been classified into three different groups. (1) Cellobiohydrolases (CBH, EC 3.2.1.91) that split β-1,4-glycosidic bonds at the non-reducing end of the cellulose molecule to yield cellobiose. (2) Endoglucanases (EG, EC 3.2.1.4) that attack internal glucosidic bonds and can also hydrolyse chemically substituted celluloses. (3) β-glucosidases that cleave short cello-oligosaccharides to glucose. Each of these three classes of enzyme contains several isoenzymes and the reason for this multiplicity of enzymes is still unclear.

Recent work has suggested that it is not possible to classify cellulases as rigidly as first proposed (Henrissat *et al.* 1985; Enari & Niku-Paavola 1987). It has been suggested that both cellobiohydrolases from *T. reesei*, CBH I and CBH II, may also function as endocellulases and that the major endocellulase EG I may be unable to attack insoluble substrates at all. Cellulolytic enzymes show synergism in the hydrolysis of crystalline cellulose.

The mechanism of the synergism shown by different pairs of cellulases (Wood & McCrae 1978; Henrissat *et al.* 1985) is still not clearly understood even though it plays an important role in cellulose hydrolysis.

The hydrolysis of cellulose is so complex that precise information concerning mechanism can only be obtained by application of recombinant-DNA technology and associated techniques such as protein engineering. For this reason, we have isolated and studied a number of different fungal genes coding for cellulase enzymes.

## Isolation of genes coding for cellulases from *T. reesei*

The gene coding for the major cellulase produced by *T. reesei*, CBH I, was isolated and characterized at VTT (Teeri *et al.* 1983) and at Cetus Corporation (Shoemaker *et al.* 1983). Since that time, we have cloned and sequenced three more genes coding for *T. reesei* cellulases, EG I (Penttilä *et al.* 1987), CBH II (Teeri *et al.* 1986) and EG III (Saloheimo, unpublished results). Full-length cDNA copies of these genes have been isolated and sequenced at VTT. In addition, a gene coding for a β-glucosidase from *Aspergillus niger* has been cloned (Penttilä *et al.* 1984) and is being characterized.

The comparison of the predicted amino acid sequences of the four *Trichoderma* cellulases reveals interesting structural features that probably have a bearing on their function. Firstly, the major cellobiohydrolase, CBH I, and the major endoglucanase, EG I, are clearly closely related, as was suggested by Bhikhabhai & Pettersson (1984) on the basis of protein sequence. The overall homology found between the amino acid sequences EG I and CBH I is about 45%. It is very likely that these two proteins have evolved from a common ancestral form to fulfil separate functions in the cellulase system. Secondly, the two cellobiohydrolases CBH I and CBH II are surprisingly different despite the similarity in their apparent enzymatic properties. Thirdly, the four cellulases analysed to date, CBH I, CBH II, EG I and EG III, all share conserved sequences that show around 70% homology to each other, as shown in figure 1. This conserved sequence is located either at the C-terminal in CBH I and EG I or at the N-terminal in CBH II and EG III. In each case, it is separated from the main part of the enzyme by a stretch of amino acids rich in serine and threonine that has been shown, at least in CBH I, to be highly glycosylated (Fägerstam *et al.* 1984).





PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

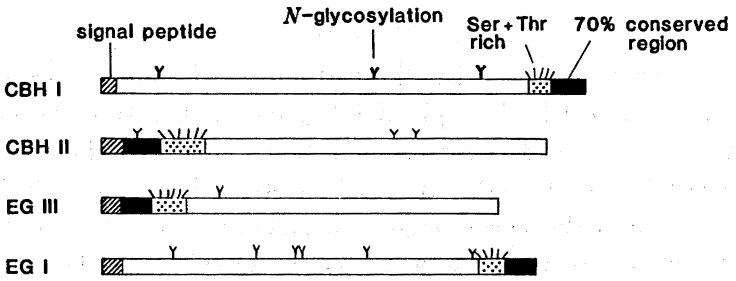


FIGURE 1.

We propose that this conserved 'domain' serves a common function or functions in all these cellulases. It is likely that this domain, which is relatively hydrophobic, is required for the binding to or the solubilization of cellulose. Experiments involving the modification of different cellulose genes and the production of mutant cellulases are under way to examine this hypothesis.

It was proposed long ago (Banks & Vernon 1963) that cellulases may act like lysozymes, which split glycosidic bonds by an acid-catalysis mechanism involving a non-ionized Glu residue as a proton donor and an ionized Asp residue as a stabilizer of the reaction intermediate (recently reviewed by Fersht (1985)). This hypothesis gained support from the demonstration that carboxyl groups were involved in the hydrolysis of β-1,4-glycosidic bonds by a fungal endoglucanase (Clarke & Yaguchi 1985). Computer-aided molecular graphics should provide a useful tool with which to develop this hypothesis further so that it can be tested by site-specific mutagenesis of the cellulase genes. Clearly in the long term, X-ray diffraction analysis of cellulase crystal structure will be required and groups both in Finland and France are working towards this goal.

## EXPRESSION OF CELLULASE GENES IN HETEROLOGOUS HOSTS

In the future it will be necessary to construct new microorganisms producing a particular limited set of enzymes for the specific modification of lignocellulose for particular purposes. In an initial attempt to produce each cellulase individually, we have transferred different cellulase cDNAs (copy DNAs) to the bacterium *Escherichia coli* and to the yeast *S. cerevisiae*. We have been able to produce CBH I fused to the N-terminal of the *cro* gene in *E. coli* such that it represented 4–5% of total cell protein. However, this hybrid protein precipitates in the cell and thus shows no enzyme activity as isolated.

To obtain expression in yeast, each of the four cellulase cDNAs was inserted into an expression vector under the control of the strong yeast promoter, phosphoglycerokinase (PGK) (Mellor *et al.* 1983). The fungal cellulase signal sequence, required for secretion from the cell, was intact in each construction. The four different recombinant yeasts, each containing a separate cellulase cDNA, were found to express and secrete the relevant cellulase into the growth medium to levels of around 1 mg $l^{-1}$. We are currently studying the properties of these recombinant cellulases, which, despite the fact that they appear to be hyperglycosylated by the yeast, are all active on the appropriate cellulosic substrates. Table 1 shows the activities of recombinant cellulases as judged by agar-plate tests.




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

TABLE 1. RECOMBINANT YEAST EXPRESSING PARTICULAR *T. reesei* CELLULASES

| cellulase cDNA transferred to yeast | substrate A | B | C | D | E |
|---|---|---|---|---|---|
| CBH I | − | − | − | − | + |
| CBH II | + | − | − | + | − |
| EG I | + | + | + | + | + |
| EG III | + | + | + | + | − |
| control | − | − | − | − | − |

Substrates: A, β-glucan; B, hydroxyethyl cellulose (HEC); C, carboxymethyl cellulose (CMC); D, lichenin; E, methylumbelliferyl cellobioside.

Properties of cellulolytic yeasts individually expressing the CBH I, CBH II, EG I and EG III cellulases from *T. reesei*.

The recombinant yeasts were grown on solid minimal media supplemented with a cellulosic substrate. The activities have been estimated from the size of the hydrolysis zone around the colony, caused by the breakdown of the substrate. The methylumbelliferyl cellobioside substrate was applied after growth of the colonies.

This work shows that new types of cellulolytic organism can be constructed that produce only specific cellulase activities; and this provides an important tool for the study of the activity of cellulases individually and in particular combinations. In addition, it provides a means of producing mutant cellulases engineered by site-specific mutagenesis for research purposes. The demonstration that yeast can secrete cellulases to the growth medium may make possible new types of organisms, perhaps even a 'cellulolytic' yeast for raising the good value of lignocellulosic wastes (Knowles *et al.* 1987).

In conclusion, to develop new processes for the utilization of lignocellulosic materials we must obtain much more knowledge concerning the biology, biochemistry and molecular biology of the organisms able to utilize this resource. The precision of recombinant-DNA technology is required for further significant progress in the study of the extremely complex lignocellulose breakdown system. There is no doubt that a great deal of effort will be required before we understand enough of the system to be able to develop novel applications utilizing lignocellulosic materials. Considerable interdisciplinary collaboration will be needed to apply this knowledge to practical goals. Once specific aims have been defined, and the enzymic steps required can be specified, novel organisms can be constructed by using recombinant-DNA technology to produce any set of enzymes needed in a particular process. It is only by understanding in detail lignocellulose and its degradation that we will be able to solve the problems that confront us both today and in the future.

## REFERENCES


Banks, B. E. C. & Vernon, C. A. 1963 *Biochem. J.* **86**, 7.

Bhikhabhai, R. & Pettersson, L. G. 1984 *FEBS Lett.* **167**, 301–308.

Clarke, A. J. & Yaguchi, M. 1985 *Eur. J. Biochem.* **149**, 233–238.

Enari, T.-M. & Niku-Paavola, M.-L. 1987 In *CRC critical reviews in biotechnology*, vol. 3. (In the press.)

Fägerstam, L. G., Pettersson, G. & Engström, J. Å. 1984 *FEBS Lett.* **167**, 309–315.

Fersht, A. 1985 *Enzyme structure and mechanism*, 2nd edn. New York: Freeman.

Henrissat, B., Driguez, H., Viet, C. & Schülein, M. 1985 *Bio/Technology* **3**, 722–726.

Knowles, J., Penttilä, M., Teeri, T., André, L. & Lehtovaara, P. 1987 In *The biochemistry and molecular biology of industrial yeasts* (ed. G. G. Stewart, I. Russell, R. D. Klein & R. R. Hiebsh), Uniscience series. Boca Raton, Florida: CRC Press. (In the press.)

Mellor, J., Dobson, M. J., Roberts, M. A., Tuite, M. F., Emtage, J. S., White, S., Lowe, P. A., Patel, T., Kingsman, A. J. & Kingsman, S. M. 1983 *Gene* **24**, 1–14.






PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Penttilä, M., Lehtovaara, P., Nevalainen, H., Bhikhabhai, R. & Knowles, J. 1987 *Gene.* (In the press.)
Penttilä, M., Nevalainen, K. M. H., Raynal, A. & Knowles, J. 1984 *Molec. gen. Genet.* **194**, 494–499.
Shoemaker, S. P., Schweickart, V., Ladner, M., Gel'fand, D., Kwok, S., Myambo, K. & Innis, M. 1983 *Bio/Technology* **1**, 691–696.
Sjöström, E. 1981 *Wood chemistry, fundamentals and applications.* New York: Academic Press.
Teeri, T., Salovuori, I. & Knowles, J. 1983 *Bio/Technology* **1**, 696–699.
Teeri, T. T., Lehtovaara, P., Kauppinen, S., Salovuori, I. & Knowles, J. 1987 *Gene.* (In the press.)
van Tilbeurgh, H., Claeyssens, M. & De Bruyne, C. K. 1982 *FEBS Lett.* **149**, 152–156.
Wood, T. M. & McCrae, S. I. 1978 *Biochem. J.* **171**, 61–71.



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES