The non-catalytic C-terminal region of endoglucanase E from *Clostridium thermocellum* contains a cellulose-binding domain

Alastair J. DURRANT,* Judith HALL,* Geoffrey P. HAZLEWOOD† and Harry J. GILBERT*‡

*Department of Agricultural Biochemistry and Nutrition, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, and †Department of Biochemistry, AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, U.K.

Mature endoglucanase E (EGE) from *Clostridium thermocellum* consists of 780 amino acid residues and has an M_r of 84016. The N-terminal 334 amino acids comprise a functional catalytic domain. Full-length EGE bound to crystalline cellulose (Avicel) but not to xylan. Bound enzyme could be eluted with distilled water. The capacity of truncated derivatives of the enzyme to bind cellulose was investigated. EGE lacking 109 C-terminal residues (EGEd) or a derivative in which residues 367–432 of the mature form of the enzyme had been deleted (EGEb), bound to Avicel, whereas EGEEa and EGEEc, which lack 416 and 246 C-terminal residues respectively, did not. The specific activity of EGEEa, consisting of the N-terminal 364 amino acids, was 4-fold higher than that of the full-length enzyme. The truncated derivative also exhibited lower affinity for the substrate β -glucan than the full-length enzyme. It is concluded that EGE contains a cellulose-binding domain, located between residues 432 and 671, that is distinct from the active site. The role of this substrate-binding domain is discussed.

INTRODUCTION

Clostridium thermocellum is an anaerobic thermophilic Gram-positive bacterium that rapidly degrades crystalline cellulose (Johnson *et al.*, 1982). Hydrolysis of the polysaccharide is mediated by a large cell-associated multienzyme complex, termed the 'cellulosome', consisting of approx. 18 polypeptides, including an apparently non-enzymic protein (S1) of M_r 210000 (Lamed & Bayer, 1988). It is proposed that S1 constitutes a cellulose-binding protein which associates with *C. thermocellum* cellulases and xylanases to form the cellulosome, although there is no definitive proof for this role (Lamed & Bayer, 1988). By using recombinant DNA methodology, numerous *C. thermocellum* endoglucanase and xylanase genes have been cloned into *Escherichia coli* (Hazlewood *et al.*, 1988). Individually, the encoded enzymes, some of which have been shown to be components of the cellulosome, have no significant activity towards crystalline cellulose (Hazlewood *et al.*, 1988). This suggests that the aggregation of enzymes to form an ordered complex, or the interaction of endoglucanases with the S1 subunit, is a prerequisite for the hydrolysis of crystalline cellulose by *C. thermocellum*. Other cellulolytic micro-organisms, notably aerobic fungi, synthesize cellulases which do not associate into large multienzyme complexes. These enzymes often contain a cellulose-binding domain (CBD) distinct from the active site (Tomme *et al.*, 1988). It is believed that, at least in fungi, the CBDs act in freeing the cellulose microfibrils in such a way that the enzyme can then perform the hydrolysis step (Tomme *et al.*, 1988). Despite extensive biochemical and ultrastructural studies of the *C. thermocellum* cellulosome, the molecular mechanisms underlying the aggregation of the cellulosome and the association of the multienzyme complex with cellulose remain unclear. Inspection of the primary structure of the cloned *C. thermocellum* cellulases which are components of the cellulosome, revealed little similarity, except for a 24-amino-acid reiterated sequence,

located generally at the C-terminus (Hall *et al.*, 1988). It has been suggested that this duplicated sequence, which is not essential for catalytic activity (Hall *et al.*, 1988; Grépinet *et al.*, 1988), interacts directly with either S1 or cellulose (Béguin, 1990).

Our research group has recently cloned and sequenced the *C. thermocellum celE* gene (Romaniec *et al.*, 1987; Hall *et al.*, 1988) encoding endoglucanase E (EGE). The mature form of the enzyme, consisting of 780 residues with an M_r of 84016, contains the conserved duplicated sequence between residues 380 and 440. The N-terminal 339 amino acids of EGE constitute a functional cellulase. The aim of the present study was to investigate the function of the C-terminal domain of the enzyme. The results show that this region constitutes a CBD which is distinct from the catalytic domain (CD). The conserved sequence is not essential for catalytic activity or cellulose binding. The putative role(s) of the CBD in EGE is discussed.

MATERIALS AND METHODS

Microbial strains, vectors and culture conditions

Escherichia coli strains employed in this study were JM83 and JM101 (Norlander *et al.*, 1983). The vectors used were pUC18, pUC19, M13mp18, M13mp19 (Norlander *et al.*, 1983), pMTL22p and pMTL23p (Chambers *et al.*, 1988). Full-length *celE* and a truncated derivative of *celE* coding for only the CD of EGE were contained in recombinant plasmids pHGB9 (Hazlewood *et al.*, 1990) and pHGB2 (Hall *et al.*, 1988) respectively. *E. coli* JM83 harbouring these plasmids was cultured in Luria broth (LB) supplemented with ampicillin (100 μ g/ml), or on LB agar containing carboxymethylcellulose (5 mg/ml); endoglucanase activity was detected by staining with Congo Red (1%, w/v; Teather & Wood, 1982). Production of β -galactosidase activity by bacterial colonies was detected by supplementing media with 5-bromo-4-chloro-3-indolyl β -D-galactopyranoside (X-gal; 4 μ g/ml).

Abbreviations used: CBD, cellulose-binding domain; CD, catalytic domain; EGE, endoglucanase E; LB, Luria broth; X-gal, 5-bromo-4-chloro-3-indolyl β -D-galactopyranoside; dATP[35 S], deoxyadenosine 5'-[α - 35 S]thio]triphosphate.

‡ To whom correspondence and reprint requests should be sent.

General recombinant DNA methodology

Agarose-gel electrophoresis, transformation of *E. coli*, plasmid preparation and the use of DNA-modifying enzymes were as described by Gilbert *et al.* (1987). DNA cloned in M13 vectors was sequenced by the dideoxy-chain-termination method (Sanger *et al.*, 1980), except that bacteriophage-T7 DNA polymerase (Tabor & Richardson, 1987) was used instead of the Klenow fragment of DNA polymerase.

Construction of *celE* derivatives

Plasmids in the pJH series were constructed as follows: truncated *celE* encoding residues 4–367 of mature EGE (essentially the CD) was excised from pHGB2 (Hall *et al.*, 1988) on a 1.1 kb *HindIII*–*SstI* restriction fragment and was cloned into M13mp19. By using site-directed mutagenesis *in vitro*, an internal *EcoRI* site at position 832 was deleted without altering the

Fig. 1. EGE derivatives encoded by pHGB9, pHGB2 and pJH1–pJH6

Deletions within EGE are indicated by Δ together with the position in the primary sequence of mature EGE of residues at the boundaries of the deletions. The N- and C-terminal amino acids of each cellulase derivative are shown. Construction of the plasmids encoding the EGE variants is described in the Materials and methods section. The positions of the restriction sites *HindIII* (H), *KpnI* (K), *NaeI* (Na), *NruI* (Nr), *PvuII* (P), *EcoRI* (R), *SaI* (Sa), *SmaI* (Sm), *SphI* (Sp), *SstI* (Ss) and *XhoI* (X) are shown. The introduction of a translational stop codon in pJH2 is depicted by 'S'. Restriction sites which have been destroyed by site-directed mutagenesis are shown in parentheses.

primary structure of the encoded enzyme and, in a second derivative, a translational stop was also created such that *celE'* encoded residues 4–364 of mature EGE (pJH2; EGEa). The two *celE* derivatives were excised on 1.1 kb *HindIII*–*EcoRI* restriction fragments and separately cloned into pMTL22p to create pJH1 (which lacks the *EcoRI* site in *celE*) and pJH2 (which lacks the *celE* *EcoRI* site and contains a translational stop codon at position 1194 of *celE*) respectively. To construct pJH4, pHGB9 (Hazlewood *et al.*, 1990) was cleaved with *EcoRI*, digested with BAL 31, blunt-ended with bacteriophage-T4 DNA polymerase and digested with *SaI*, which cleaves downstream of *celE*. The blunt-ended *SaI* DNA fragment of approx. 1.1 kb (encoding the 3' region of *celE*) was cloned into pJH1 which had been digested with *EcoRI* and *XhoI*. Nucleotide sequencing of selected plasmids revealed that the 5' region of the reconstituted *celE* gene was out of frame with the 3' end of the gene. Consequently, hybrid *celE* was digested with *EcoRI*, filled in with sequenase, and re-ligated. To verify that the resultant *celE* gene now constituted a single open reading frame (encoded by pJH4), the relevant region of *celE* was resequenced, and the cell-free extract from *E. coli* cells harbouring pJH4 was shown to contain a thermostable (67 °C for 1 h) protein of M_r 77 000 that was absent from strains containing either pMTL22p alone or the plasmids harbouring the *celE* derivatives in which the 5' and 3' regions of the gene were out of phase (results not shown). Plasmid pJH6 was constructed by cloning the 1.95 kb *SphI*–*KpnI* restriction fragment from pHGB9 (Hazlewood *et al.*, 1990) into pMTL22p. Finally, pJH5 was synthesized by digesting pJH6 with *SstI* and re-ligating the 4.2 kb fragment generated. The structures of the *celE* derivatives are shown in Fig. 1.

Purification of EGE

E. coli JM83 harbouring truncated *celE* cloned in pMTL22p was cultured for 16 h in LB (1 litre) containing ampicillin (100 μ g/ml). A periplasmic fraction which contained 70 % of the total endoglucanase activity of the culture was prepared from the harvested cells as described by Hazlewood *et al.* (1990), and was heated at 60 °C for 10 min with gentle swirling. Denatured protein was removed by centrifuging (15 000 g ; 4 °C; 20 min). Endoglucanase activity in the supernatant was precipitated with $(\text{NH}_4)_2\text{SO}_4$ (48 % saturation), redissolved in 50 mM-Tris/HCl, pH 8.0, and dialysed against three changes of 10 mM-Tris/HCl, pH 8.0. EGE contained in the dialysed periplasmic extract was further purified by hydrophobic interaction chromatography on a Phenyl-Sepharose column (2.6 cm $\times$ 12.0 cm), essentially as described by Chauvaux *et al.* (1990). Column fractions were monitored by SDS/PAGE (Laemmli, 1970), and pure EGE was eluted with 8 % (v/v) ethylene glycol.

Cellulose binding studies

Cell-free extract (5 ml), prepared by sonication from *E. coli* cells (100 ml of culture) harbouring full-length or truncated forms of *celE*, was mixed with 5 ml of Avicel (PH105; FMC Corp., Philadelphia, PA, U.S.A.; 5 %, w/v) suspended in 100 mM-Tris/HCl, pH 7.5 (Buffer A). After shaking gently for 1 h at 0 or 60 °C, Avicel was recovered by filtration, washed three times with Buffer A, and bound protein was eluted with 1 ml of distilled water. The M_r of proteins eluted from Avicel was determined by SDS/PAGE, using 10 % (w/v) acrylamide gels as described by Laemmli (1970).

Assays

The preparation of cell-free extracts and the assay of endoglucanase activity were as described by Romaniec *et al.* (1987).

Reducing sugar was determined by the dinitrosalicylic acid method of Miller (1959), and was expressed as glucose equivalents. Protein was measured as described by Lowry *et al.* (1951), using BSA as standard.

Enzymes and reagents

Restriction enzymes, bacteriophage-T4 DNA ligase and polymerase were from Bethesda Research Laboratories. Bacteriophage-T7 DNA polymerase and other reagents for DNA sequencing were from Cambridge Bioscience. The deoxyadenosine 5'-[α - 35 S]thiotriphosphate (dATP[35 S]; 1000 mCi/mmol) was supplied by Amersham International. BAL 31 was obtained from Boehringer. Unless specified, all other reagents were supplied by Sigma.

RESULTS

Binding to cellulose of full-length and truncated forms of EGE

Previous studies showed that *celE* encodes an endoglucanase (EGE), of M_r 84000 that exhibits some xylanase activity. The C-terminal region of the protein was not essential for catalytic activity (Hall *et al.*, 1988). To determine the function of this non-catalytic domain of the enzyme, the capacity of full-length EGE and a truncated derivative of the endoglucanase lacking the C-terminal 416 residues (EGEa; Fig. 1) to bind crystalline cellulose (Avicel) was investigated. Results (Fig. 2) showed that cell-free extracts from *E. coli* harbouring pHGB9, which encodes full-length EGE, expressed a protein of M_r 84000 which bound to both Avicel and amorphous cellulose at 0 °C, but not to xylan. The protein could be eluted with distilled water, exhibited endoglucanase activity at 60 °C and was precipitated with anti-EGE antiserum. When the binding experiments were repeated at 60 °C, similar results were obtained. By contrast, *E. coli* cells

Fig. 2. Binding of EGE to polysaccharides

Cell-free extracts of *E. coli* cells harbouring pHGB9 and pJH2 are shown in lanes 1 and 6 respectively. Protein eluted with distilled water from *E. coli* protein/Avicel mixtures, incubated at 0 °C, is shown in lanes 2 (pHGB9) and 7 (pJH2). Lane 3 contains enzyme released from a pHGB9 cell-free extract/Avicel suspension incubated at 60 °C. Bound proteins eluted with distilled water from mixtures containing *E. coli* (harbouring pHGB9) proteins incubated with amorphous cellulose and xylan, at 0 °C, are shown in lanes 4 and 5 respectively.

Fig. 3. Binding of truncated derivatives of EGE to Avicel

Enzyme eluted from Avicel with distilled water, and proteins present in cell-free extracts, are shown in lanes 1 and 2 respectively. Proteins were prepared from *E. coli* cells harbouring pHGB9 (a), pJH6 (b), pJH4 (c) and pJH5 (d).

harbouring pJH2, which encodes EGEa, failed to synthesize a protein which bound to insoluble cellulose or xylan. Similarly, purified EGEa did not interact with either polysaccharide (results not shown).

To locate more precisely the position of the CBD in EGE, derivatives of *celE* encoding truncated forms of the enzyme were constructed as described in the Materials and methods section (Fig. 1). The capacity of the EGE derivatives to bind cellulose was investigated. The data (Fig. 3) revealed that pJH6, which encodes EGEb, a truncated form of EGE lacking the 110 C-terminal residues, directed the synthesis of a protein of M_r 72000

Fig. 4. Activity of native and truncated forms of EGE against β -glucan

Native EGE (○) and the truncated derivatives EGEb (●) and EGEa (■) were assayed for β -glucanase activity at different substrate concentrations.

which bound to Avicel, could be eluted from the polysaccharide with distilled water and hydrolysed carboxymethylcellulose at 60 °C. EGEc, a derivative of EGE from which the 246 C-terminal amino acids had been deleted, did not interact with Avicel. EGEb, in which residues 367–432 of mature EGE had been deleted, bound to crystalline cellulose and could be eluted from the polysaccharide in an active form with distilled water. EGEb contains only the eight C-terminal residues of the second 24-amino-acid duplicated sequence which is highly conserved in other *C. thermocellum* endoglucanases (Hall *et al.*, 1988). These data indicate that the region of EGE which confers the ability to bind to cellulose is located between residues 432 and 671 and is distinct from the active site. Furthermore, the conserved sequence does not play an important role in either cellulose binding or catalytic activity.

Kinetic properties of native and truncated derivatives of EGE

Previous studies showed that EGE is very active against barley β -glucan. The role of the CBD in the hydrolysis of this substrate by EGE was investigated. The data (Fig. 4) revealed that EGEa, which lacks the CBD, had a specific activity of 75.4 units/mg of protein, compared with 12.96 units/mg of protein for EGE synthesized by *E. coli* harbouring the entire *celE* gene. The full-length endoglucanase displayed typical Michaelis–Menten kinetics, with an apparent K_m of 1.2 mg/ml for β -glucan. A similar result was obtained for the enzyme EGEb, which did not contain the conserved sequence, but retained the CBD. By contrast, EGEa, which lacks the CBD, did not exhibit Michaelis–Menten kinetics at low substrate concentration, where very low activity was observed. These data suggest that the CBD plays a role in catalysis only at low substrate concentration.

DISCUSSION

It is now widely accepted that the multienzyme cellulase complex of *C. thermocellum*, termed the cellulosome, is composed of 14–18 endoglucanases and includes one major subunit (S1) which is apparently devoid of enzyme activity. It has been suggested (Lamed & Bayer, 1988) that this subunit is a cellulose-binding protein which functions in anchoring the other components of the cellulosome and accounts for the high affinity of the complex for crystalline cellulose. At least one study (Wu *et al.*, 1988) has demonstrated that a component with properties very similar to S1 acts in concert with a smaller endoglucanase to slowly hydrolyse Avicel. If indeed the S1 subunit mediates the attachment of other cellulosome components to cellulose, there is no obvious requirement for the individual endoglucanases of the cellulosome to contain CBDs distinct from the active site. In the present study we have shown that EGE, which is a component of the cellulosome (Hazlewood *et al.*, 1990), contains a CBD that is distinct from the catalytic domain. Our data suggest that the binding of the cellulosome to cellulose is not the exclusive property of a single component, such as the S1 subunit, but results from the sum of the affinities for cellulose of individual enzymes like EGE. At present it is unclear how many of the cellulosome components have CBDs, but in this context, Grépinet *et al.* (1988) showed that the 508 N-terminal residues of xylanase Z (another proven component of the cellulosome) are not essential for catalytic activity. By analogy with EGE, it would seem entirely possible that the N-terminal domain of xylanase Z constitutes a substrate-binding domain, even though it shows no sequence similarity to the CBD of EGE.

The present study shows that deletion of the reiterated 24-amino-acid sequence, conserved in other *C. thermocellum* endoglucanases and xylanase Z, does not affect the interaction of EGE

with cellulose. This argues against the notion, suggested by Grépinet & Béguin (1986), that the conserved sequence could be a CBD.

An increase in cellulase specific activity against soluble substrates, such as we observed on removing the CBD from EGE, has been previously reported for an endoglucanase from *Thermomonospora fusca* (Ghangas & Wilson, 1988), but was not apparent for cellulases from *Trichoderma reesei* (Tomme *et al.*, 1988) or *Bacteroides succinogenes* (McGavin & Forsberg, 1989). *In vivo*, there is some evidence that EGE exists both as a full-length and a truncated form (M_r 43000) which lacks the CBD (Hazlewood *et al.*, 1990). Knowles *et al.* (1988) have suggested that the proteolytic removal of carbohydrate-binding domains from cellulases and other polysaccharide-degrading enzymes, such as glucoamylase, may be an important mechanism for the modification of their substrate specificities during the course of hydrolysis of complex insoluble substrates. Once the substrate has been degraded to short chain-length soluble oligosaccharides, enzymes with reduced affinity for the original highly polymeric substrate, but with enhanced activity against the soluble breakdown products, would be available. Micro-organisms can thus increase the diversity of enzymes expressed while maintaining a limited number of genes. Alternatively, it could be argued that the primary role of EGE in cellulolysis is to hydrolyse soluble substrates. Thus, early in cellulolysis, when the concentration of soluble substrate (cellulo-oligosaccharides) is low, possession of a CBD distinct from the active site improves the affinity of EGE for its substrate. As digestion proceeds, increased levels of soluble cellulose-oligosaccharides are released and EGE no longer requires the CBD, which is removed through proteolysis, resulting in higher activity against the depolymerized cellulose and shorter-chain intermediates.

We thank Keith Davidson for skilled technical assistance and the Agricultural and Food Research Council (grant LRG 138) for supporting this work.

REFERENCES

- Béguin, P. (1990) Annu. Rev. Microbiol., in the press
- Chambers, S. P., Prior, S. E., Barstow, D. A. & Minton, N. P. (1988) *Gene* **68**, 139–149
- Chauvaux, S., Béguin, P., Aubert, J.-P., Bhat, K. M., Gow, L. A., Wood, T. M. & Bairoch, A. (1990) *Biochem. J.* **265**, 261–265
- Ghangas, G. S. & Wilson, D. B. (1988) *Appl. Environ. Microbiol.* **54**, 2521–2526
- Gilbert, H. J., Jenkins, G., Sullivan, D. A. & Hall, J. (1987) *Mol. Gen. Genet.* **210**, 551–556
- Grépinet, O. & Béguin, P. (1986) *Nucleic Acids Res.* **14**, 1791–1799
- Grépinet, O., Chebrou, M.-C. & Béguin, P. (1988) *J. Bacteriol.* **170**, 4582–4588
- Hall, J., Hazlewood, G. P., Barker, P. J. & Gilbert, H. J. (1988) *Gene* **69**, 29–38
- Hazlewood, G. P., Romaniec, M. P. M., Davidson, K., Grépinet, O., Béguin, P., Millet, J., Raynaud, O. & Aubert, J.-P. (1988) *FEMS Microbiol. Lett.* **51**, 231–236
- Hazlewood, G. P., Davidson, K., Clarke, J. H., Durrant, A. J., Hall, J. & Gilbert, H. J. (1990) *Enzyme Microb. Technol.* **12**, 656–662
- Johnson, E. A., Sakajoh, M., Halliwell, G., Madia, A. & Demain, A. L. (1982) *Appl. Environ. Microbiol.* **43**, 1125–1132
- Knowles, J., Teeri, T. T., Lehtovaara, P., Penttilä, M. & Saloheimo, M. (1988) in *Biochemistry and Genetics of Cellulose Degradation* (Aubert, J.-P., Béguin, P. & Millet, J., eds.), pp. 153–169. Academic Press, London and San Diego
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lamed, R. & Bayer, E. A. (1988) *Adv. Appl. Microbiol.* **33**, 1–46
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275
- McGavin, M. & Forsberg, C. W. (1989) *J. Bacteriol.* **171**, 3310–3315
- Miller, G. L. (1959) *Anal. Chem.* **31**, 426–428

- Norrander, J., Kemp, T. & Messing, J. (1983) *Gene* **26**, 101–106
- Romaniec, M. P. M., Clarke, N. G. & Hazlewood, G. P. (1987) *J. Gen. Microbiol.* **133**, 1297–1307
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178
- Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4767–4771
- Teather, R. M. & Wood, P. J. (1982) *Appl. Environ. Microbiol.* **43**, 777–780
- Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vande-kerckhove, J., Knowles, J., Teeri, T. & Claeyssens, M. (1988) *Eur. J. Biochem.* **170**, 575–581
- Wu, J. H. D., Orme-Johnson, W. H. & Demain, A. L. (1988) *Biochemistry* **27**, 1703–1709

Received 26 June 1990/9 August 1990; accepted 14 August 1990