Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase

Luis M. A. FERREIRA,* 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.

Xylanase A (XYLA) from *Pseudomonas fluorescens* subspecies *cellulosa* shows sequence conservation with two endoglucanases from the same organism. The conserved sequence in XYLA, consisting of the *N*-terminal 234 residues, is not essential for catalytic activity. Full-length XYLA and a fusion enzyme, consisting of the *N*-terminal 100 residues of XYLA linked to mature alkaline phosphatase, bound tightly to crystalline cellulose (Avicel), but not to xylan. The capacity of truncated derivatives of the xylanase to bind polysaccharides was investigated. XYLA lacking the first 13 *N*-terminal animo acids did not bind to cellulose. However, a catalytically active XYLA derivative (XYLA′), in which residues 100–234 were deleted, bound tightly to Avicel. Substrate specificity, cellulose-binding capacity, specific activity and K_m for xylan hydrolysis were evaluated for each of the xylanases. No differences in any of these parameters were detected for the two enzymes. It is concluded that XYLA contains a cellulose-binding domain consisting of the *N*-terminal 100 residues which is distinct from the active site. Spatial separation of the catalytic and cellulose-binding domains is not essential for the enzyme to function normally.

INTRODUCTION

Hydrolysis of the plant structural polysaccharides cellulose and hemicellulose is effected by microbial enzymes. Endo-β-1,4-glucanase (EC 3.2.1.4), exo-β-1,4-glucanase (EC 3.2.1.91) and β-glucosidase (EC 3.2.1.21) act co-operatively to degrade cellulose, whereas xylan, the major polymeric component of hemicellulose, is hydrolysed by β-1,4-xylanase (EC 3.2.1.8) and β-xylosidase (EC 3.2.1.37) action. Multiple forms of cellulase and xylanase produced by prokaryotic species can result from post-transitional processing of a single gene product (Biely, 1985), but are, more often, the products of multigene families (Hazlewood *et al.*, 1988; Gilbert *et al.*, 1987, 1988). Enzymes encoded by the cellulase gene families of a number of cellulolytic bacteria including *Fibrofactor succinogenes* (McGavin & Forsberg, 1989), *Cellulomonas fimi* (Ong *et al.*, 1989) and *Thermomonospora fusca* (Ghangas & Wilson, 1988) contain two distinct functional domains, a catalytic domain (CD) and a cellulose-binding domain (CBD). Exo-β-1,4-glucanase from *Trichoderma reesei* appears to require a CBD for efficient hydrolysis of crystalline cellulose (Tomme *et al.*, 1988; Van Tilbeurgh *et al.*, 1986), but the role of these domains in bacterial cellulolysis remains to be elucidated.

Our previous work has focused on the isolation and characterization of multiple endoglucanase and xylanase genes from *Pseudomonas fluorescens* subsp. *cellulosa* (Gilbert *et al.*, 1987, 1988, 1990; Hall & Gilbert, 1988; Hall *et al.*, 1989). The nucleotide sequences of two endoglucanase genes (*celA* and *celB*) and a xylanase gene (*xynA*) have been determined, and the primary sequence of the encoded proteins and their biological properties deduced. The *N*-terminal regions of endoglucanase B (EGB) and xylanase A (XYLA) showed sequence identity with the *C*-terminus of endoglucanase A (EGA). In all three enzymes, the conserved regions contained serine-rich sequences, whose function is at present unknown, but is not related to catalytic activity. We have recently shown that a CBD is located within the conserved region of EGB (Gilbert *et al.*, 1990).

The purpose of the present study was to determine whether the conserved sequence in XYLA also constitutes a CBD which is independent of the CD, and to elucidate the role of the serine-rich sequences in enzyme function. Our results show for the first time that a xylanase contains a CBD, which in XYLA is located within the first 100 *N*-terminal residues; the remainder of the conserved sequence, including the serine-rich regions, plays no detectable role in XYLA function. Evidence is also presented which suggests that the spatial separation of the CBD and CD, by the serine-rich linker sequences, is not essential for the two domains to perform their respective functions. A possible role for the linker sequences in domain shuffling is discussed.

MATERIALS AND METHODS

Microbial strains, vectors and culture conditions

Escherichia coli strains employed in this study were JM83, JM101 (Norrander *et al.*, 1983) and the *phoA* mutant CC118 (Manoil & Beckwith, 1985). The vectors used were pUC19, M13mp18 and M13mp19 (Norrander *et al.*, 1983). Full-length *xynA* and a truncated derivative of the gene were contained in recombinant plasmids pRS6 and pRS16 respectively (Hall *et al.*, 1989). *E. coli* strains were cultured in Luria broth (LB) or on LB/agar. Recombinant phage were grown in 2 × yeast/tryptone medium. Ampicillin (100 μg/ml) was used to select for transformants. Functional xylanase and alkaline phosphatase respectively were detected by the addition of the chromogenic substrates 4-*O*-methyl-*O*-glucorono-D-xylan–Remazol Brilliant Blue R (RBB) (3 μg/ml; Sigma Chemical Co.) and 5-bromo-4-chloro-3-indolyl phosphate (2 μg/ml) to media.

Abbreviations used: CBD, cellulose-binding domain; CD, catalytic domain; EGA, endoglucanase A; EGB, endoglucanase B; XYLA, xylanase A; LB, Luria broth; dATP[^{35}S], deoxyadenosine 5′-[α-[^{35}S]thio]triphosphate; RBB, Remazol Brilliant Blue R.

‡ To whom correspondence and reprint requests should be sent.

General recombinant-DNA procedure

Agarose-gel electrophoresis, transformation of *E. coli*, plasmid isolation 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, essentially as described by Sanger *et al.* (1980).

Cellulose-binding studies and purification of XYLA

The periplasmic fraction (5 ml; Hsiung *et al.*, 1986) prepared from *E. coli* cells (100 ml of culture), harbouring full-length or truncated forms of *xynA*, 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 8.0 (Buffer A). After shaking gently for 1 h at 0 °C, Avicel was separated from residual unbound protein by filtration, and the crystalline cellulose was washed a further three times with Buffer A. Enzyme bound to the polysaccharide was eluted either in 1 ml of 10% (w/v) SDS with heating to 100 °C for 5 min or in 5 ml of 6 M-guanidinium chloride. XYLA activity was recovered after dialysing material eluted with guanidinium chloride against 2 × 1000 vol. of Buffer A at 4 °C. The M_r of proteins eluted from Avicel was determined by on 10% (w/v) acrylamide gels as described by Laemmli (1970).

Assays

Xylanase activity was measured by the method of Biely *et al.* (1985). A unit of enzyme activity released 1 μmol of xylose equivalent/min. Protein was determined as described by Lowry *et al.* (1951), with BSA as standard.

Enzymes and reagents

All restriction enzymes and phage-T_4 DNA ligase were from Bethesda Research Laboratories. Phage-T_7 DNA polymerase and other reagents for DNA sequencing were from Cambridge Biosciences. The deoxyadenosine 5′-[α-[^{35}S]thio]triphosphate (dATP[^{35}S]) (1000 mCi/μmol) was supplied by Amersham International. Phage-T_4 DNA polymerase was purchased from Pharmacia, and *Bal*31 was obtained from Boehringer. Unless specified, all other reagents were supplied by Sigma.

Large-scale preparations of XYLA were made by affinity chromatography on Avicel. Periplasmic proteins, prepared from *E. coli* cultures (800 ml), were incubated with 15 ml of 5% (w/v) Avicel in Buffer A. After washing the cellulose, bound protein was eluted with 10 ml of 6 M-guanidinium chloride and was dialysed twice against 1000 vol. of Buffer A. A truncated form of XYLA, containing only the CD, was purified from *E. coli* harbouring pRS16, as described by Hall *et al.* (1989).

Construction of *xynA* derivatives

Plasmids pLH1–pLH6 were constructed as follows. Full-length *xynA*, in pRS6 (Hall *et al.*, 1989), was digested with restriction endonuclease *Eco*RV, which cleaves *xynA* 73 bp downstream of the serine-rich coding region (Hall *et al.*, 1989). The plasmid was then cut with *Bal*31, blunt-ended with phage-T_4 DNA polymerase and digested at an *Eco*RI site which is downstream of the protein-coding region of *xynA* in the polylinker region of pUC19. The region of *xynA* encoding only the CD was excised from pRS16 (Hall *et al.*, 1989). on a *Eco*RI/blunt-ended restriction fragment, and ligated into *Bal*31-treated pRS6. The resultant recombinant plasmids were transformed into *E. coli* JM83, and xylanolytic colonies were selected on RBB–xylan. The *xynA–phoA* hybrid was constructed in a similar way; pRS6 digested with *Bal*31 as described above was ligated to the *phoA* gene, which had been isolated on a 1.25 kb *Stu*I–*Hin*dIII fragment coding for residue 8 (of the mature enzyme) to the *C*-terminus of alkaline phosphatase, which had been blunt-ended by treatment with phage-T_4 DNA polymerase. Recombinants were selected in a *phoA*$^-$ mutant of *E. coli*. The plasmid pRS8 was constructed in a previous study (Hall *et al.*, 1989). The extent of the deletions within *xynA* were determined by nucleotide sequencing of appropriate regions of the gene cloned in M13 vectors.

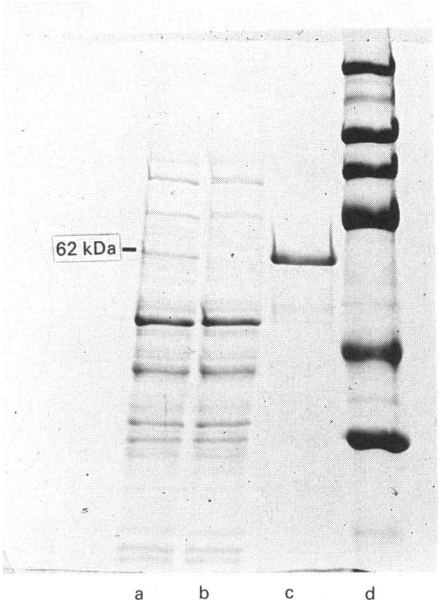

Fig. 1. Binding of XYLA to crystalline cellulose (Avicel)

Periplasmic proteins from *E. coli* cells harbouring pRS6 (contains the complete *xynA* gene) were incubated with Avicel, and the bound protein was eluted with SDS. The tracks contained the following material: a, total periplasmic proteins from RS6; b, protein from strain RS6, which did not bind Avicel; c, protein eluted from Avicel; d, M_r standards from RS6.

RESULTS AND DISCUSSION

Binding of XYLA to cellulose

The mature form of XYLA from *P. fluorescens* subsp. *cellulosa* encoded by *xynA* has a M_r of 62000 (Hall *et al.*, 1989). Previous studies have shown that the *N*-terminal portion of XYLA contains two regions rich in serine residues and exhibits sequence conservation with the *N*- and *C*-terminal domains of *P. fluorescens* endoglucanase A (EGA) and endoglucanase B (EGB) respectively (Gilbert *et al.*, 1990). Gene-deletion studies showed the conserved regions not to be essential for catalytic activity (Hall *et al.*, 1989) and, in the case of EGB, to contain a CBD (Gilbert *et al.*, 1990). To determine whether the comparable *N*-terminal sequence in XYLA also binds cellulose, total periplasmic proteins from *E. coli* harbouring pRS6, a recombinant plasmid containing the entire *xynA* gene (Hall *et al.*, 1989), were incubated with crystalline cellulose (Avicel). Results (Fig. 1) showed that a 62000-M_r protein bound specifically to Avicel and could be eluted with 6 M-guanidinium chloride or 10% (w/v) SDS. Protein bound to Avicel or eluted with guanidinium chloride exhibited xylanase activity, confirming its identity as XYLA. A truncated derivative of XYLA encoded by pRS16, which lacks the *N*-terminal 234 residues of the mature form of the enzyme, including the conserved domain, did not bind to crystalline cellulose. This indicates that the *N*-terminal domain of the xylanase, which shows sequence conservation with the *N*- and *C*-termini of EGB and EGA respectively, constitutes a CBD. However, it could be argued that the capacity of XYLA to bind cellulose depends on an interaction between the *N*-terminal region and other sequences within the enzyme. To investigate this possibility, the *N*-terminal

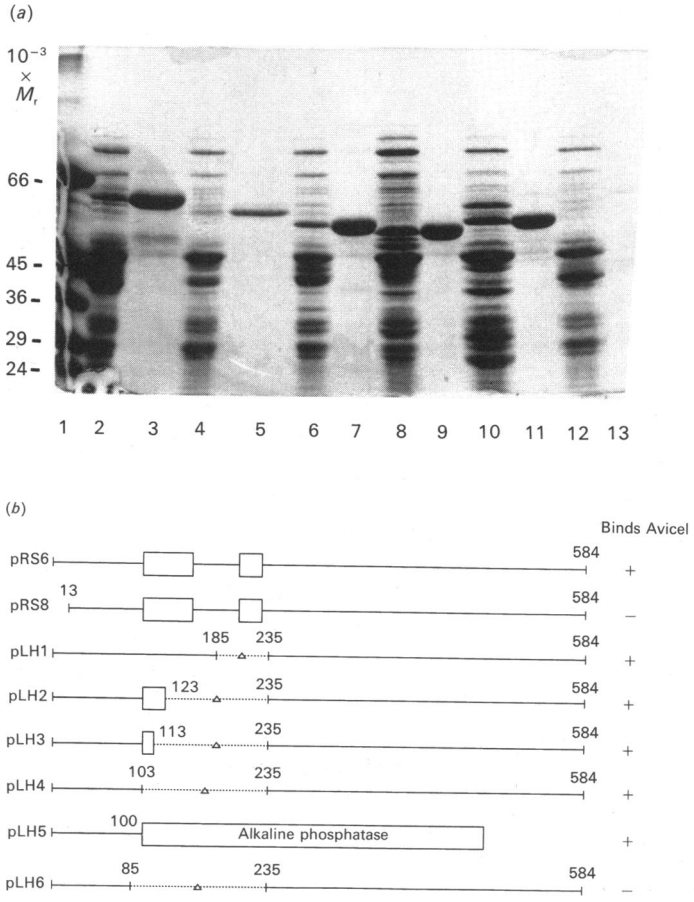

Fig. 2. Binding of truncated derivatives of XYLA to Avicel

(*a*) SDS/PAGE of proteins present in the periplasm (lanes 2, 4, 6, 8, 10 and 12), and recovered by boiling in 10% (w/v) SDS after binding to Avicel (lanes 3, 5, 7, 9, 11 and 13). Track 1 contained M_r standards. Periplasmic proteins were prepared from *E. coli* strains harbouring pRS6 (lanes 2 and 3), pLH1 (lanes 4 and 5), pLH3 (lanes 6 and 7), pLH4 (lanes 8 and 9), pLH5 (lanes 10 and 11) and pLH6 (lanes 12 and 13). (*b*) XYLA derivatives encoded by pRS6, pRS8 and pLH1–pLH6. The two serine-rich sequences are depicted as boxes. Deletions within XYLA are indicated by △ together with the position in the primary sequence of residues at the boundaries of the deletions. Periplasmic proteins were prepared from *E. coli* strains harbouring modified forms of *xynA*. Binding to Avicel was evaluated as described in the Materials and methods section.

coding region of *xynA* was fused in-frame to the *E. coli phoA* gene (encodes alkaline phosphatase). The fusion protein expressed in *E. coli* bound tightly to cellulose, whereas mature alkaline phosphatase did not (Fig. 2). These data confirm that the *N*-terminal domain of XYLA constitutes a CBD that functions independently of the rest of the enzyme. To assess whether the CBD of XYLA also bound to xylan, full-length XYLA and the XYLA–alkaline phosphatase fusion protein were incubated separately with insoluble xylan. Data revealed no interaction between the proteins and polysaccharide, indicating that the CBD does not also act as a xylan-binding domain.

The precise role of the CBDs now known to be present in some bacterial cellulases remains to be elucidated. The present study shows, for the first time, that such domains also occur in a xylanase which does not hydrolyse $\beta(1\rightarrow4)$ glycosidic linkages between glucose residues in a variety of cellulose molecules. This argues against the notion that possession of a CBD confers the ability to hydrolyse cellulose (West *et al.*, 1989). The relevance of a CBD to an enzyme which hydrolyses xylan, but not cellulose,

Table 1. Binding of XYLA derivatives to crystalline cellulose (Avicel)

Plasmid	Residues deleted	Predicted M_r of XYLA	Avicel binding	M_r of protein bound to Avicel	Xylanase activity
pRS6	0	62300	+	61100	+
pRS8	1–13	60800	–	NP*	+
pLH1	185–235	56900	+	55200	+
pLH2	123–235	51300	+	53600	+
pLH3	113–235	50400	+	51800	+
pLH4	103–235	49300	+	50000	+
pLH6	85–235	47400	–	NP*	+

* Abbreviation: NP, no protein detected.

is an important question. In natural ecosystems, cellulolytic bacteria frequently occur in close proximity to decaying plant tissue. Our experiments have shown that the CBD in XYLA is not directly involved in xylan hydrolysis *in vitro*, but it seems likely that, in a normal habitat, possession of a CBD would confer a selective advantage in allowing the xylanase to remain in intimate contact with plant cell walls, which contain both cellulose and hemicellulose.

Binding of XYLA derivatives to Avicel

To examine the precise location of the CBD within the *N*-terminal region of XYLA, a series of 5′ and 3′ truncated derivatives of the CBD coding region of *xynA* were constructed and fused in-frame to the CD of XYLA, encoded by pRS16. The capacity of the modified xylanases to bind cellulose was investigated. Results (Fig. 2 and Table 1) showed that removal of only 13 residues from the *N*-terminus of mature XYLA resulted in the loss of cellulose binding. *E. coli* strains harbouring truncated *xynA* derivatives in which residues 103–235, 123–235, 113–235, 185–235 of XYLA had been deleted, produced proteins of M_r 50000, 51800, 53600, and 55000 respectively, all of which bound to cellulose. Each of these enzymes exhibited xylanase activity, confirming their identity as XYLA derivatives. These data indicate that the CBD is located within the first 100 *N*-terminal residues of XYLA. The remaining 135 amino acid residues of the conserved region, which include both serine-rich domains, play no role in cellulose binding or catalytic activity. To verify this conclusion the substrate specificity, specific activity, cellulose-binding capacity and K_m for hydrolysis of soluble xylan were determined for the complete enzyme, a truncated XYLA consisting of the CD only, and a XYLA derivative lacking the serine-rich sequences (XYLA′). Results (Table 2) revealed no differences between the three forms of the enzyme, confirming that the serine-rich linker regions do not have an important role in enzyme function or cellulose binding. This argues against the widely held view that the spatial separation of discrete domains in cellulases is crucial for correct enzyme function (Gilkes *et al.*, 1990).

Recent studies have revealed that prokaryotic and eukaryotic cellulases are composed of discrete domains which display substantial conservation of sequence (Gilbert *et al.*, 1990; Gilkes *et al.*, 1990; Teeri *et al.*, 1987). These domains are linked together by sequences, highly enriched in hydroxy amino acids, which are often extensively *O*-glycosylated (Fägerstam *et al.*, 1985). This post-translational modification affords cellulases protection from proteolysis (Langsford *et al.*, 1987). The position of these conserved domains varies between enzymes (Warren *et al.*, 1986; Hall & Gilbert, 1988; Gilbert *et al.*, 1990), indicating that cellulases evolved through the acquisition and shuffling of common ancestral sequences encoding discrete domains.

Table 2. Properties of XYLA encoded by pRS6, pRS16 and pLH4

Plasmid	Specific activity (units/mg of protein)*	Relative activity: Soluble xylan	Relative activity: Insoluble† xylan	Relative activity: 4-*O*-Methyl-glucurono-D-xylan†	Avicel binding (50% binding)‡	K_m§
pLH4	330	1	0.63	0.47	2.4	1.19
pRS16	324	1.1	0.65	0.48	0	1.17
pRS6	339	1.1	0.69	0.57	2.9	1.16

* Specific activity against soluble xylan for enzyme purified to homogeneity.
† Activities expressed relative to activity of pLH4 XYLA against soluble xylan, which was designated as 1.0.
‡ Amount of Avicel which will bind 50% of 0.145 unit of XYLA under the conditions described in Fig. 1.
§ Measured for the hydrolysis of soluble xylan.

In eukaryotes, functional domains of proteins are often encoded by distinct exons (Doolittle, 1979). Extended non-coding sequences (introns) provide the means by which these exons can by excised, without interruption of coding sequence, and fused to other exons to generate novel proteins. The shuffling of functional domains within bacterial cellulases and xylanases points to the existence of a similar mechanism. The DNA sequences encoding serine-rich linker sequences in cellulases and xylanases may fulfil a role analogous to that of introns. In the absence of these linker sequences, precise excision of sequences encoding the functional domains would be required, to avoid interruption of coding sequence essential for catalytic function or cellulose binding. Thus the serine-rich encoding regions of *xynA* and other cellulase genes contain extended sequences (the two serine-rich sequences in XYLA are encoded by 228 bp), which can be cleaved to allow excision of CD- or CBD-encoding regions and their subsequent fusion to genes which would then direct the expression of a novel enzyme. The serine residues would be glycosylated and thus provide protection from proteolytic cleavage of the linker sequences.

We thank the Agricultural and Food Research Council (grant LRG 138) for supporting this work.

REFERENCES

Biely, P. (1985) Trends Biotechnol. **3**, 286–290
Biely, P., Mislovicova, D. & Toman, R. (1985) Anal. Biochem. **144**, 142–146
Doolittle, R. F. (1979) in The Proteins (Neurath, H. & Hill, R. L., eds.), vol. 4, pp. 1–118, Academic Press, New York
Fägerstam, L. G., Pettersson, L. G. & Engstrom, J. A. (1985) FEBS Lett. **167**, 309–315
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
Gilbert, H. J. Sullivan, D. A., Jenkins, G., Kellett, L. E., Minton, N. P. & Hall, J. (1988) J. Gen. Microbiol. **134**, 3239–3247
Gilbert, H. J., Hall, J., Hazlewood, G. P. & Ferreira, L. M. A. (1990) Mol. Microbiol., in the press
Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. J. (1990) J. Biol. Chem., in the press
Hall, J. & Gilbert, H. J. (1988) Mol. Gen. Genet. **213**, 112–117
Hall, J., Hazlewood, G. P., Huskisson, N. S., Durrant, A. J. & Gilbert, H. J. (1989) Mol. Microbiol. **3**, 1211–1219
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
Hsiung, H. M., Mayne, N. G. & Becker, G. W. (1986) Bio/Technology **4**, 991–995
Laemmli, U. K. (1970) Nature (London) **227**, 680–685
Langsford, M. L., Gilkes, N. R., Singh, B., Moser, B., Miller, R. C., Jr., Kilburn, D. G. & Warren, R. A. (1987) FEBS Lett. **225**, 163–167
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
Manoil, C. & Beckwith, J. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 8129–8133
Norrander, J., Kemp, T. & Messing, J. (1983) Gene **26**, 101–106
Ong, E., Greenwood, J. M., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. J. (1989) Trends Biotechnol. **7**, 239–243
Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) J. Mol. Biol. **143**, 161–178
Teeri, T. T., Lehtovaara, P., Kauppinen, S., Salovuori, I. & Knowles, J. (1987) Gene **51**, 43–52
Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vanderkerckhore, J., Knowles, J., Teeri, T. T. & Claeyssens, M. (1988) Eur. J. Biochem. **170**, 575–581
Van Tilbeurgh, H., Tomme, P., Claeyssons, M., Bhikhabhai, R. & Pattersson, G. (1986) FEBS Lett. **204**, 223–227
Warren, R. A. J., Beck, C. F., Gilkes, N. R., Kilburn, D. G., Langsford, M. L., Miller, R. C., Jr., O'Neill, G. P., Scheufens, M. & Wong, W. K. R. (1986) Proteins **1**, 335–341
West, C. A., Elzanowski, A., Yeh, L.-S. & Barker, W. (1989) FEMS Microbiol. Lett. **59**, 167–172

Received 27 February 1990/26 April 1990; accepted 2 May 1990