Biochem. J. (1990) 271, 277–280 (Printed in Great Britain)

The tertiary structure of a bacterial cellulase determined by small-angle X-ray-scattering analysis

Ingrid PILZ,* Erika SCHWARZ,* Douglas G. KILBURN,† Robert C. MILLER, Jr.,† R. Antony J. WARREN† and Neil R. GILKES†‡

*Institut für Physikalische Chemie der Universität Graz, Heinrichstrasse 28, A-8010 Graz, Austria, and ‡Department of Microbiology, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada

CenA from *Cellulomonas fimi* is a β-1,4-endoglucanase that binds tightly to cellulose. X-ray-scattering analyses show that the enzyme is tadpole-shaped: the previously identified catalytic and cellulose-binding domains comprise the head and tail respectively. It appears that this structural and functional organization is common to several cellulases from bacteria and fungi.

INTRODUCTION

The extracellular enzyme systems produced by micro-organisms to degrade cellulose contain multiple components. Cellulases from the soft-rot fungus *Trichoderma reesei* are the most extensively studied; several bacterial cellulases have also been closely examined. Many such enzymes bind strongly to cellulose. The mechanism and significance of this interaction remain unknown, but it has recently been demonstrated that certain such β-1,4-glucanases, including CBH I and CBH II from *T. reesei* and CenA and Cex from the cellulolytic bacterium *Cellulomonas fimi*, are bifunctional proteins [1–5]. They comprise structurally and functionally independent catalytic and cellulose-binding domains (CBD) joined by a linker sequence or 'hinge' region rich in hydroxy amino acids. In CBH I the *C*-terminal A block (the CBD) is joined to the catalytic domain or 'core' by the *O*-glycosylated B block hinge region. In CBH II the arrangement of the domains is reversed: the *N*-terminal A block is joined to the catalytic core by a duplicated B block sequence [1]. CenA and Cex, also glycoproteins [6], are similarly organized [3,4]. In CenA the *N*-terminal CBD is joined by a 23-amino acid-residue linker containing only proline and threonine residues (the Pro–Thr box) to its catalytic domain. In Cex this arrangement is reversed. The functional domains of all these enzymes can be separated by proteolytic cleavage at the junction of the catalytic domain [1,4,5].

SAXS analysis has shown that CBH I and CBH II are each shaped like a tadpole: the A and B blocks form the extended, perhaps flexible, tail, and the core peptide (catalytic domain) the ellipsoidal head [7,9]. Differences in the respective tail structures were attributed to the duplication of the B block in CBH II. The present paper describes a model for the structure of CenA determined by SAXS analysis of the non-glycosylated form of the enzyme synthesized in *Escherichia coli* from recombinant *C. fimi* DNA [10]. This molecule is also tadpole-shaped, but has features that distinguish it from the fungal enzymes.

EXPERIMENTAL

Materials

ngCenA was purified from *E. coli* JM101 (pUC18–1.6*cenA*) as described previously [4] and transferred to 50 mM-potassium phosphate, pH 7.0 (phosphate buffer), containing 0.02% NaN_3 by ultrafiltration on a PM10 membrane (Amicon). p30 (the isolated ngCenA catalytic domain or core) was prepared by incubation of 130 mg of ngCenA in 10.0 ml of phosphate buffer containing 440 units of crude *C. fimi* proteinase and 0.02% NaN_3 at 37 °C for 48 h. Digestion was monitored by SDS/PAGE as described previously [4]. The reaction was stopped by addition of 6 mg of phenylmethanesulphonyl fluoride dissolved in 300 μl of anhydrous acetone, and the volume was reduced to 1.5 ml by concentration on a PM10 membrane. Final purification was by gel-filtration chromatography on a Superose 12 column (Pharmacia LKB Biotechnology) equilibrated with phosphate buffer. The final yield after concentration was 71 mg of p30 (= 80% of theoretical maximum). Analytical ultracentrifugation showed that the ngCenA and p30 samples were homodisperse and both were judged to be at least 99% homogeneous by SDS/PAGE (results not shown.).

Absorption coefficients ($A^{0.1\%}_{1\,cm,\,205}$) for ngCenA and p30 (32.78 and 35.12 respectively) were determined by the far-u.v. method of Scopes [11]. These values were used to calculate the absolute concentrations of solutions of these proteins, which were then used to determine absorption coefficients at 280 nm. These latter values (2.64 and 2.38 respectively) agreed well with the predicted values (2.46 and 2.41 respectively) calculated from the tyrosine and tryptophan content of the protein or peptide and its predicted M_r [12].

Small-angle X-ray-scattering analyses

Experimental scattering curves [$I(h)$ functions] were recorded stepwise with a conventional Kratky camera in the range of h = 0.094 to 4.47 nm^{-1} [$h = (4\pi \cdot \sin\theta)/\lambda$, where 2θ = the scattering angle and λ = 0.154 nm (wavelength of Cu K_α-line used)]. The samples were cooled to 4 °C during the scattering experiments. A series of five concentrations was measured for both intact ngCenA and p30. The concentration of the intact enzyme varied between 13 and 54 mg/ml, and that of p30 between 12 and 84 mg/ml. The scattering at each concentration was recorded several times at 115 different scattering angles; at least 200000 pulses were counted per point. The samples were investigated in phosphate buffer at pH 7.0 with (ngCenA) or without (p30) 0.02% NaN_3. Data evaluation, desmearing and indirect Fourier transformation were done as described previously [13–15].

Abbreviations used: CBH I and CBH II, *Trichoderma reesei* β-1,4-glucan cellobiohydrolases I and II respectively; CenA, *Cellulomonas fimi* β-1,4-endoglucanase A; Cex, *C. fimi* β-1,4-exoglucanase (previous nomenclature of CenA and Cex are explained in ref. [4]; the prefix ng indicates the non-glycosylated form synthesized in *Escherichia coli* from recombinant DNA); CBD, cellulose-binding domain; SAXS, small-angle X-ray scattering; R_g, radius of gyration; $D_{max.}$, maximum dimension.

‡ To whom correspondence should be addressed.

Model calculations

Theoretical scattering [$I(h)$] and electron-pair distance-distribution [$p(r)$] functions were calculated by using a finite-element method. Models were composed of small identical spheres with a diameter smaller than the resolution of the SAXS method. The theoretical scattering curves of the models were calculated by using Debye's formula, and the distance-distribution functions were calculated independently from the co-ordinates of the spheres, i.e. without Fourier transformation of the scattering curve [16]. The curves derived from over 100 different models for intact ngCenA were tested for correspondence to the experimental data. These included structures in which the head and tail regions were arranged co-axially or were fixed at other angles. Models were also tested in which a range of angles was varied to simulate movement.

RESULTS

Intact ngCenA

The following structural parameters were derived from the scattering curve [$I(h)$ function] and the electron-pair distance distribution [$p(r)$ function] (Figs. 1 and 2 respectively): $R_g = 5.35 \pm 0.05$ nm; $D_{max.} = 21.0 \pm 1.0$ nm (Table 1). The $I(h)$ and $p(r)$ functions for ngCenA are similar to those obtained for CBH I and CBH II [7–9]. The models previously calculated for CBH I and CBH II are tadpole-shaped. The data indicated a similar structure for ngCenA, but the mass of the ngCenA tail is clearly greater because the frequency of the distances between 5 and 10 nm [the maxima in the $p(r)$ function; Fig. 2] is much higher than in the $p(r)$ functions for CBH I and CBH II [7–9]. In the structural models for CBH I and CBH II, the ellipsoidal head and extended tail share a common longitudinal axis. None of the many similar models calculated for ngCenA satisfied the experimental data. However, the $p(r)$ function computed for a model comprising a truncated ellipsoidal head and elongated tail in which the respective longitudinal axes were constrained at an angle of 135° (Fig. 2 inset) fitted the experimental data well (Fig. 2). The R_g values for ngCenA and CBH II are almost identical,

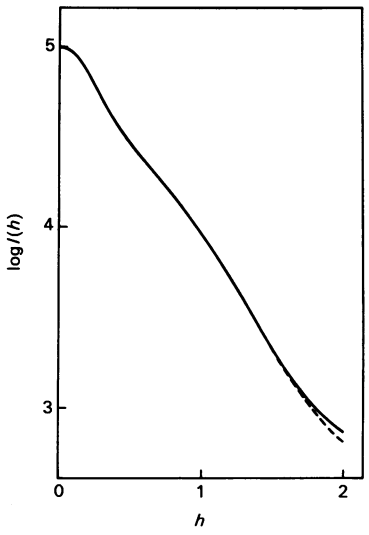

Fig. 1. Comparison of the experimental scattering curve [$I(h)$ function] for ngCenA (——) and the calculated curve (----) for the model shown in Fig. 2 inset

$I(h)$ = scattering intensity (normalized to 10^5 at $h = 0$); $h = 4\pi \cdot \sin\theta/\lambda$ (nm^{-1}), where 2θ = scattering angle and $\lambda = 0.154$ nm.

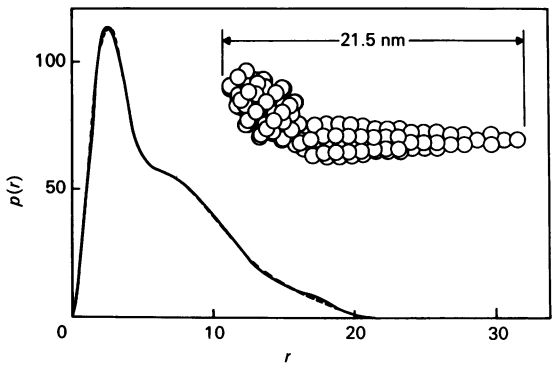

Fig. 2. Comparison of the experimental electron-pair distance-distribution [$p(r)$] function for ngCenA (——) with calculated error bars and the calculated $p(r)$ function (----) for the model shown in the inset

$p(r)$ values (arbitrary units) are normalized to the same maximum value for both cures; r = intramolecular distance in real space (nm).

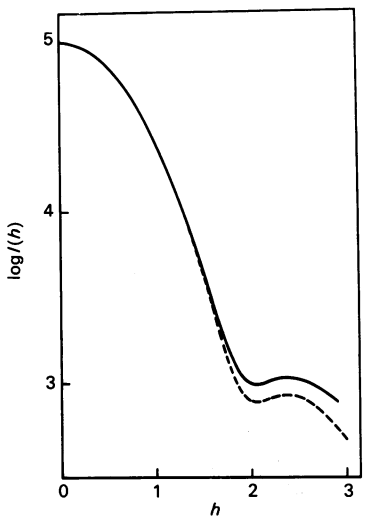

Fig. 3. Comparison of the experimental scattering curve [$I(h)$ function] for the ngCenA core peptide (p30) (——) and the calculated curve (----) for the model shown in Fig. 4 inset

$I(h)$ and h are defined in the legend to Fig. 1.

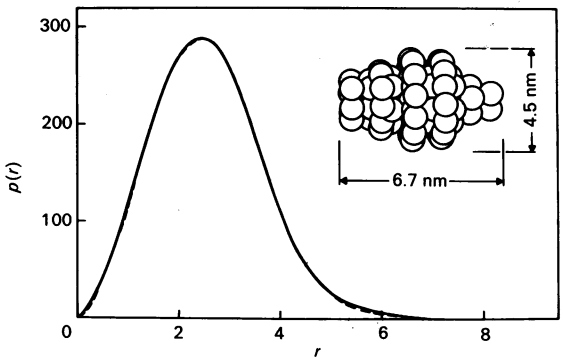

Fig. 4. Comparison of the electron-pair distance-distribution [$p(r)$ function for the ngCenA core peptide (p30) with calculated error bars (——) and the calculated $p(r)$ function (----) for the model shown in the inset.

$p(r)$ and r are defined in the legend to Fig. 2.

Table 1. Molecular parameters for ngCenA from *C. fimi* and CBH I and CBH II from *T. reesei* determined by small-angle X-ray-scattering analysis

Values (±S.E.M.) are given for the intact enzymes and their core peptides (isolated catalytic domains).

	Intact enzyme			Core peptide		
	ngCenA	CBH I	CBH II	ngCenA	CBH I	CBH II
R_g (nm)	5.35±0.05	4.27±0.2	5.4±0.1	1.98±0.03	2.09±0.2	2.1±0.1
$D_{max.}$ (nm)	21.0±1.0	18.0±0.5	21.5±0.5	6.4±0.3	6.5±0.3	6.0±0.3

as are the $D_{max.}$ values, but significantly greater than the corresponding parameters for CBH I (Table 1).

ngCenA core peptide

The $I(h)$ function and corresponding $p(r)$ function for the ngCenA core peptide (p30) are shown in Figs. 3 and 4 respectively. The structure of the core peptide could be modelled as a truncated ellipsoid with axes $a = 6.7$ nm and $b = c = 4.5$ nm (Fig. 4 inset). During the refinement of the model, a better fit with the experimental data was obtained for a model in which the electron density was not homogeneous but was lower towards the centre of the particle (Fig. 4). As in the corresponding models for CBH I and CBH II, the core corresponds to the head of the tadpole-shaped intact enzyme. The molecular parameters determined for the ellipsoidal cores of all three enzymes (Table 1) are very similar.

DISCUSSION

Previous biochemical analyses have shown that ngCenA is a bifunctional protein comprising independent catalytic and cellulose-binding domains joined by a short linker sequence, the Pro–Thr box [3–6]. SAXS analysis has now shown that ngCenA is tadpole-shaped: it comprises an ellipsoidal head or core (catalytic domain) and an extended tail (Pro–Thr box plus CBD) (Figs. 2 and 4 insets). The gross structural and functional organization of ngCenA (418 amino acid residues) thus closely resembles those of CBH I and CBH II (497 and 447 amino acid residues respectively), two cellulolytic enzymes from the fungus *T. reesei* [1,7,9]. The model for ngCenA incorporates a constrained angle of 135° between the long axes of the core and tail regions. This is in contrast with the models obtained for CBH I and CBH II, where co-axial arrangements fitted the experimental data. Co-axial models were unsatisfactory for ngCenA, as were models in which a range of angles between the core and tail regions were averaged to simulate movement.

The core peptides (isolated catalytic domains) of ngCenA, CBH I and CBH II contain widely different numbers of amino acid residues (284, 431 and 365 respectively). SAXS analyses indicate that all three comprise ellipsoidal structures of similar size (Table 1). The lower electron density indicated for the centre of the ngCenA core is represented in the model by the looser arrangement of the spheres that simulate protein mass. The ngCenA tail (CBD plus Pro–Thr box) comprises 134 amino acid residues and its length, determined from the model (Fig. 2 inset), is 15.8 nm. SAXS analysis has shown that the CBH II tail (A, B and B′ blocks; 82 amino acid residues) is of comparable length (15.5 nm); the CBH I tail (A and B blocks; 66 amino acid residues) is shorter (12.9 nm). However, ngCenA has a significantly greater protein mass in the tail region than has either of the two *T. reesei* enzymes. CBH I and CBH II are glycoproteins (7–11% carbohydrate). Most of the glycosyl substituents are *O*-linked to hydroxy amino acids of the B or B′ blocks [17] and probably account for part of the mass of the tail structures determined by SAXS analysis [9]. CenA isolated from *C. fimi* is also glycosylated [6], containing approx. 8% carbohydrate (M. L. Langsford, N. R. Gilkes, R. C. Miller, Jr., D. G. Kilburn & R. A. J. Warren, unpublished work), but the enzyme used in the present study was the more readily prepared non-glycosylated form synthesized in *E. coli* from recombinant *C. fimi* DNA [2]. Glycosylation of CenA does not appear to alter catalytic activity or binding to cellulose, but it has been shown to stabilize the enzyme against proteolysis [2], and we have speculated that it involves *O*-glycosylation of threonine residues within the Pro–Thr box [2]. However, the nature and position of the glycosyl substituents and their contribution to the tertiary structure of the enzyme remain to be determined.

Although the structural and functional organizations of ngCenA, CBH I and CBH II are similar, there are no obvious similarities between the primary structures of the CenA CBD [10] and the A blocks (i.e. binding domains) of the fungal enzymes [18]. However, several other bacterial β-1,4-glucanases contain terminal domains with significant sequence similarity to the CenA CBD. For example, Cex from *C. fimi* contains a 108-amino acid-residue *C*-terminal domain that also mediates binding to cellulose [4]; a similar 104-amino acid-residue domain occurs at the *N*-terminus of MbcelA from *Microbispora bispora* [19]. These terminal domains are characterized in particular by conserved cysteine residues near their *N*- and *C*-extremities and by four conserved tryptophan residues (see Fig. 5 in ref. [5]) and are joined to their respective catalytic domains by linker sequences closely related to the Pro–Thr box [3,19]. Moreover, the catalytic domains of Cex and MbcelA (315 and 326 amino acid residues respectively) have similar M_r values to that of the ngCenA core. On the basis of these considerations it is likely that Cex and MbCelA also adopt a tadpole conformation in solution. Other bacterial β-1,4-glucanases, for example EGA from *Pseudomonas fluorescens* subsp. *cellulosa* [20,21] and CenB from *C. fimi* (A. Meinke, N. R. Gilkes, R. C. Miller, Jr., D. G. Kilburn & R. A. J. Warren, unpublished work), also contain terminal domains with sequence similarity to the CenA CBD, but the adjoining catalytic domains (approx. 860 and 940 amino acid residues respectively) have considerably greater M_r values, which may indicate a somewhat different tertiary structure.

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Österreichischer Fonds zur Förderung der Wissenschaftlichen Forschung. We thank Emily Kwan and William Craig for expert technical assistance and Dr. Hermann Esterbauer for helpful discussions.

REFERENCES

1. Tomme, P., van Tilbeurgh, H., Pettersson, G., van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. & Claeyssens, M. (1988) Eur. J. Biochem **170**, 575–581
2. Langsford, M. L., Gilkes, N. R., Singh, B., Moser, B., Miller, R. C., Jr., Warren, R. A. J. & Kilburn, D. G. (1987) FEBS Lett. **225**, 163–167

3. 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
4. Gilkes, N. R., Warren, R. A., Miller, R. C., Jr. & Kilburn, D. G. (1988) J. Biol. Chem. **263**, 10401–10407
5. Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. J. (1989) J. Biol. Chem. **264**, 17802–17808
6. Gilkes, N. R., Langsford, M. L., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. J. (1984) J. Biol. Chem. **259**, 10455–10459
7. Schmuck, M., Pilz, I., Hayn, M. & Esterbauer, H. (1986) Biotechnol. Lett. **8**, 397–402
8. Abuja, P. M., Pilz, I., Claeyssens, M. & Tomme, P. (1988) Biochem. Biophys. Res. Commun. **156**, 180–185
9. Abuja, P. M., Schmuck, M., Pilz, I., Tomme, P., Claeyssens, M. & Esterbauer, H. (1988) Eur. J. Biophys. **15**, 339–342
10. Wong, W. K. R., Gerhard, B., Guo, Z. M., Kilburn, D. G., Warren, R. A. J. & Miller, R. C., Jr. (1986) Gene **44**, 315–324
11. Scopes, R. K. (1974) Anal. Biochem. **59**, 277–282
12. Cantor, C. R. & Schimmel, P. R. (1980) Biophysical Chemistry, vol. 2, pp. 380–381, W. H. Freeman, San Francisco
13. Glatter, O. & Kratky, O. (eds.) (1982) Small Angle X-Ray Scattering, Academic Press, New York
14. Pilz, I., Glatter, O. & Kratky, O. (1979) Methods Enzymol. **61**, 148–249
15. Glatter, O. (1977) J. Appl. Crystallogr. **10**, 415–421
16. Glatter, O. (1980) Acta Phys. Austriaca **52**, 234–256
17. Fägerstam, L. G., Pettersson, L. G. & Engström, J. A. (1984) FEBS Lett. **167**, 309–315
18. Teeri, T. T., Lehtovaara, P., Kauppinen, S., Salovuori, I. & Knowles, J. (1987) Gene **51**, 43–52
19. Yablonsky, M. D., Bartley, T., Elliston, K. O., Kahrs, S. K., Shalita, Z. P. & Eveleigh, D. E. (1988) in Biochemistry and Genetics of Cellulose Degradation (Aubert, J.-P. Béguin, P. & Millet, J., eds.), pp. 249–266, Academic Press, London
20. Hall, J. & Gilbert, H. J. (1988) Mol. Gen. Genet. **213**, 112–117
21. Gilbert, H. J., Hall, J., Hazlewood, G. P. & Ferreira, L. M. A. (1990) Mol. Microbiol. **4**, 759–767

Received 17 May 1990/24 July 1990; accepted 1 August 1990