Dynamic light scattering study of the two-domain structure of *Humicola insolens* endoglucanase V

Claire Boisset[a], Redouane Borsali[a], Martin Schülein[b], Bernard Henrissat[a,*]

[a]*Centre de Recherches sur les Macromolécules Végétales, CNRS, BP 53, F-38041 Grenoble cedex 9, France*
[b]*Novo-Nordisk als, Novo Allé, DK-2880 Bagsvaerd, Denmark*

Received 15 September 1995; revised version received 20 October 1995

Abstract Endoglucanase V (EG V) of *Humicola insolens* is composed of a catalytic domain and of a cellulose-binding domain linked by a 33 amino acid long peptide rich in Ser, Thr and Pro residues. This work describes the dynamic behavior of the two-domain structure of EG V as revealed by quasi-elastic light scattering experiments. For both the full-length and the isolated catalytic domain, the autocorrelation function is essentially described by a single relaxation mode. The equivalent hydrodynamic radius of the catalytic domain was found to correspond precisely to the dimensions measured from the previously determined three-dimensional structure. The results obtained with the full-length protein allow a description of the two domain structure of EG V similar to that resulting from earlier studies using small angle X-ray scattering on cellulases from *Trichoderma reesei*. The hydrodynamic dimensions of the entire enzyme can be approximated as an ellipsoid with dimensions of 42 × 133.6 Å.

Key words: Cellulase; *Humicola insolens*; Endoglucanase; Domain structure; Dynamic light scattering

1. Introduction

Cellulases are often made of two domains, a catalytic domain linked to a cellulose-binding domain by a hinge region of variable length. The two-domain structure of cellulases was first demonstrated with *Trichoderma reesei* and *Cellulomonas fimi* enzymes [1,2] and such an architecture was later found to occur also in xylanases and other plant cell-wall degrading enzymes [3,4]. It has long been noted that removal of the cellulose-binding domain from a two-domain cellulase dramatically reduces the activity of the truncated enzyme on microcrystalline cellulose [1,2,5] and this has led to suggest that the spatial arrangement of the two constitutive domains would be essential for optimal activity. The three-dimensional structures of a number of isolated cellulase catalytic domains [6–12] and of cellulose-binding domains [13,14] have been determined. However, it has so far proven impossible to crystallize an entire cellulase, i.e one containing the two domains. This could be due to the possible flexibility of the linker peptide, or to its heterogeneous glycosylation, or to both. According to these considerations, studies aiming at elucidating the spatial relationship between the individual domains are particularly important. Small angle X-ray scattering experiments have been performed on several entire cellulases and their isolated catalytic domains and models were built as to best fit the experimental data. The best fitting models of two-domain cellulases featured a tadpole shape, while catalytic domains were satisfactorily modelled as ellipsoids [15–17]. These very useful models do not describe the possible flexibility of the domains about the hinge region nor account for the highly heterogeneous *O*-glycosylation of the linker peptide. Here, a dynamic light scattering study of endoglucanase V (EG V) from *Humicola insolens* reports the hydrodynamic dimensions of the entire two-domain enzyme as well as of the isolated catalytic domain.

2. Experimental

2.1. Preparation of EG V and EG V-core

EG V from *H. insolens* was cloned and expressed in *Aspergillus oryzae*. The full sequence (284 amino acids) and the construction was described by Rasmussen et al. [18]. The cloned product was recovered after fermentation by separation of the extracellular fluid from the production organism. EG V was then highly purified by affinity chromatography using 150 g of Avicel in 20 mM sodium phosphate buffer, pH 7.5. The slurry was mixed with the crude fermentation broth which contained a total of about 1 g of EG V protein. After mixing at 4°C for 20 min, the Avicel–enzyme mixture was packed into a column with a dimension of 50 × 200 mm. The column was washed with 200 ml buffer (20 mM sodium phosphate, pH 7.5), then washed with 0.5 M NaCl in the same buffer until no more protein was eluted. The column was then washed with 500 ml of 20 mM Tris buffer, pH 8.5. Finally, the pure full-length enzyme was eluted with 1% triethylamine at pH 11.8.

The eluted enzyme solution was adjusted to pH 8.0 and concentrated to more than 5 mg protein/ml using an Amicon cell unit equipped with a Dow GR61PP membrane (polypropylene with a cut-off of 20 kDa). The enzyme has a molecular mass of 43 kDa in SDS-PAGE. Using mass spectrometry, a mass around 37 kDa was obtained with some heterogeneity due to different amount of *O*-glycosylation on the linker (between 22 and 40 sugar units per molecule). Mature EG V consists of a 213 residues core, a 33 amino acid linker containing 22 Thr and Ser residues all of which are *O*-glycosylated, and a *C*-terminal 38 amino acid long cellulose-binding domain (CBD) which is homologous to that published in [13].

The catalytic core domain (EG V-core) was obtained by introducing a stop codon after the nucleotide which codes for residue 213 in the mature enzyme. The new construction has been transformed into *A. oryzae*. The transformed strain was fermented as above and the extracellular proteins recovered free from the production organism. EG V-core was purified by cation exchange chromatography. The fermentation broth was adjusted to pH 3.5 and filtrated free of precipitated proteins. Then the proteins were ultra-filtrated (concentrated and washed with water) on a Dow GR81PP membrane with a cut-off 6 kDa until the conductivity was below 1 mS/cm. The sample was finally applied to a S-Sepharose column equilibrated with 20 mM citrate buffer, pH 3.5. The enzyme bound to S-Sepharose at this low pH and was eluted as a single peak using a sodium chloride gradient from 0 to 0.5 M. The eluded pure enzyme was concentrated on a Amicon cell with the Dow GR81PP membrane. The truncated protein has a pI of 5.0 and a molecular mass of 22 kDa in both SDS-PAGE and electro spray mass spectrometry, in agreement with the amino acid composition and indicating no glycosylation.

*Corresponding author.

SSDI 0014-5793(95)01244-3

2.2. Dynamic light scattering

2.2.1. Samples preparation. The sample solutions were prepared with sodium acetate buffer (50 mM, pH 5.0) made with de-ionized and Millipore Alpha-Q filtered water. To eliminate dust and other large particles, all samples were further filtered on 0.1 μm polyvinylidene difluoride filters (Millipore) prior to measurements. Concentration losses upon filtering were estimated by UV absorption and found to be approximately 4% and corrected.

2.2.2. Equipment and data analysis. The quasi-elastic scattering measurements were performed at the temperature 25 ± 0.1°C using a ALV (Langen, Germany) apparatus equipped with an automatic goniometer table, a digital ratemeter and a temperature controlled sample cell. The scattered light of a vertically polarized λ_o=4880 Å argon laser (Spectra-Physics model 2020, 3 W, operating at around 0.3 W) was measured at different angles in the range of 70–150° corresponding to $1.96 \times 10^{-3} < q(Å^{-1}) < 3.3 \times 10^{-3}$ where $q = (4\pi n/\lambda_o)\sin(\theta/2)$; θ is the scattering angle, and n is the refractive index of the medium (n = 1.33). The full homodyne autocorrelation functions of the scattered intensity, measured in steps of 10° in the scattering angle, were obtained using the ALV-5000 autocorrelator. The intermediate dynamic scattering function $I(q,t)$ is related to the measured homodyne intensity–intensity time correlation function by the Siegert-relation [19]:

$$G^{(2)}(q,t) = B\,[1 + \alpha|I(q,t)|^2] \quad (1)$$

where B is the baseline and α is the spatial coherence factor which depends on the geometry of the detection and the ratio of the intensity scattered by the particle to that scattered by the solvent. For a Brownian motion, the autocorrelation function is generally described by a single relaxation, i.e. $I(q,t) \sim e^{-\Gamma t}$. Γ is the relaxation frequency ($1/\tau$) and is related to the diffusion coefficient D by the relation $\Gamma = Dq^2$. The autocorrelation function of the scattered intensity was analyzed by means of the cumulant method to yield the effective diffusion coefficient as a function of the scattering angle. Additionally, the constrained regularization method (CONTIN) developed by Provencher [20] was used to obtain the distribution $A(\tau)$ of decay times. For the latter, a statistical parameter 'probability to reject' is calculated for each solution, and the suggested one is that closest to 0.5.

$$\left[\frac{G^{(2)}(q,t)}{B} - 1\right]^{1/2} = \int_0^\infty A(\tau)e^{-(t/\tau)}\,d\tau \quad (2)$$

These methods are now routinely used to analyze the quasi-elastic light scattering data for many systems and allow the determination of the relaxation modes which characterize the dynamic behavior of macromolecules.

3. Results and discussion

The experiments were performed at angles between 70° and 150°, for several protein concentrations comprised between 0.2 and 1.12 mg/ml. One should note, however, that the scattering angle θ = 70° is by no means the minimum reachable angle in these experiments. Measurements were carried out at angles of 70° and above, to minimize the possible influence of dust particles that may have survived the filtration process. A typical autocorrelation function obtained for the full-length EG V protein at a concentration c = 1 mg/ml and scattering angle θ = 90° is displayed in Fig. 1. The dots represent the experimental data, and the solid line the monoexponential fit. This figure shows that the autocorrelation function is essentially represented by a single exponential decay. However, a minor contribution from a slow process can be noticed at high decay times (Fig. 1). The CONTIN analysis confirmed the existence of this minor relaxation mode which can be attributed to aggregates and that could easily be separated from the dynamics of the protein. A similar observation was also made with EG V-core. The formation of aggregates is beyond the scope of this paper. Fig. 2 represents the variation of the fast relaxation frequencies $\Gamma(q)$ attributed to the unaggregated proteins as a function of q^2 at the concentration of 0.5 mg/ml for EG V and EG V-core. The angular variation of these frequencies $\Gamma = 1/\tau_s$ measured for both systems shows a q^2 behavior indicating a diffusive motion [21].

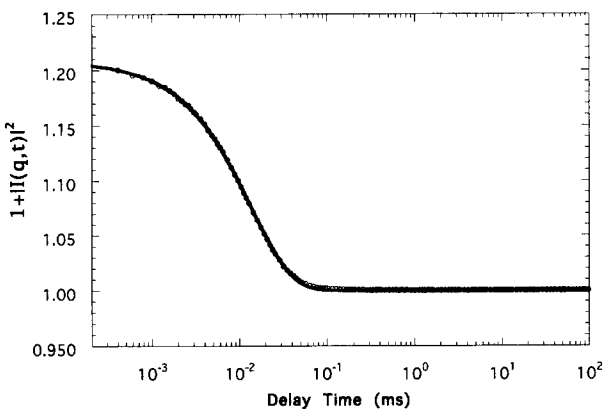

Fig. 1. Typical autocorrelation function as measured by quasi-elastic light scattering on EG V (1 mg/ml) at 25°C at θ = 90°. The dots represent the experimental data and the solid line is a monoexponential fit.

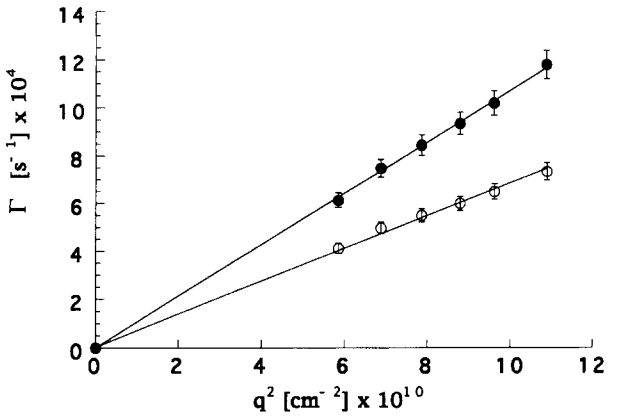

Fig. 2. Variation of the relaxation frequencies Γ as a function of q^2 for EG V (open circles) and EG V-core (filled circles) at a concentration of 0.5 mg/ml.

In the range of concentrations that were used, the initial slopes of the curves $\Gamma(q)$ versus q^2 yield the translational diffusion coefficients $D_T(c) = (\Gamma_T/q^2)_{q\to 0}$. Their variation as a function of the concentration is displayed in Fig. 3. Within experimental errors, D_T was found independent of the concentration for full-length EG V, while a slightly negative value of the slope was found with EG V-core, indicating a lower thermodynamic

Table 1
Dimensions of the constitutive elements of three fungal cellulases

Enzyme (source)	EG V (*H. insolens*)	CBH I (*T. reesei*)	CBH II (*T. reesei*)
CBD (residues)	38	36	36
Linker (residues)	33	23	40
Core (residues)	213	435	364
Hydrodynamic diameter of the core	42 Å	65 Å	60 Å
Estimated length of the full enzyme	133.6 Å	180 Å	210 Å
Reference	this work	[17]	[16]

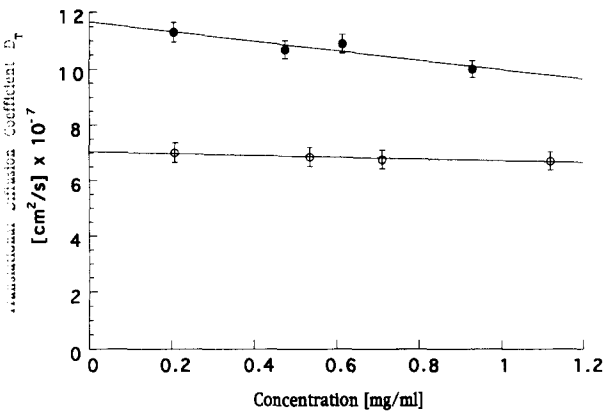

Fig. 3. Variation of the translational diffusion coefficient D_T as a function of concentration of EG V (open circles) and EG V-core (filled circles).

stability for the truncated protein. It is worth noting that the negative value of the slope is qualitatively in good agreement with the presence of a few aggregates. To go further in our analysis, the $D_T^0 = D_T(c)_{c\to 0}$ were converted to the equivalent hydrodynamic radii (R_H) using the Stokes-Einstein relation $R_H = k_B T/(6\pi\eta_s D_T^0)$ where k_B is the Boltzmann's constant, η_s the viscosity of the solvent and T the absolute temperature. This gave hydrodynamic radii of 34.7 Å and 21 Å for the full-length and the core enzyme, respectively, assuming a spherical shape for both particles. The radius obtained for EG V-core (21 Å) is in excellent agreement with the three-dimensional structure which displays a flattened spheroidal shape with dimensions of 42 × 42 × 22 Å [8] and illustrate the precision of the technique.

With its modular structure, the entire EG V protein cannot be described as a simple sphere. A model made of two spheres (corresponding to the catalytic domain and the CBD) at a distance from each other was not selected because the transversal size of the highly glycosylated linker cannot be neglected compared to that of the small CBD. Because of this, a two-domain modular enzyme such as EG V is almost certainly better described as a prolate ellipsoid whose dimensions can be derived from the translational diffusion coefficent by the D_T^0 by the relations [22]:

$$D_T^0 = \left(\frac{kT}{6\pi\eta_s a}\right) G(\rho) \quad (3)$$

with

$$G(\rho) = (1-\rho^2)^{-1/2} \ln\left(\frac{1+(1-\rho^2)^{1/2}}{\rho}\right) \quad (4)$$

where ρ is the axial ratio ($\rho = b/a$), a and b the lengths of the major and minor semi-axes, respectively. Assuming that the minor semi-axis of the ellipsoid is approximately equal to the hydrodynamic radius of the core, i.e. b = 21 Å, a value of a = 66.8 Å is found, giving an overall extension of 133.6 Å for the full-length enzyme. Although the measurement of the diffusion coefficient is very precise, the calculated extension of the protein depends, of course, on the accuracy of the model and the potential error is probably larger than that made on the size of the core.

Table 1 summarizes the dimensions of the various constitutive elements of *H. insolens* EG V along with those of two previously studied enzymes from *T. reesei*, cellobiohydrolases I and II (CBH I and CBH II) [16,17]. The three enzymes carry homologous CBDs of almost certainly identical dimensions. When the two cellobiohydrolases are compared, one can immediately see that, since the dimension of the core of CBH II is smaller than that of CBH I by 5 Å, a 35 Å difference is generated by the 17 extra residues in the linker of CBH II compared to that of CBH I (Table 1). The resulting average value of 2.05 Å per residue in the linker is different from the 1.5, 3.1 and 3.5 Å per residue found in α-helices, polyproline helices and β-strands, respectively. On the other hand, this value is close to the 2.0 Å per residue found in 3_{10} helices and is very similar to the value of 2.1 Å found for *C. hordeins* of barley, a Pro/Gln-rich mixture of proteins [22].

When an average value of ~2 Å per residue is applied to the 33 aa linker of EG V, a length of ~66 Å is found. When the dimensions of the core (42 Å) and of the linker (66 Å) are substracted from the dimension of the full-length EG V, a dimension of 26 Å is obtained for the CBD. This value is in excellent agreement with the dimensions of the three-dimensional structure of the CBD [13]. The case of the two cellobiohydrolases is not as clear since a value of ~2 Å per residue would result in linker dimensions of 46 and 80 Å for CBH I and CBH II, respectively. These values, added to the dimensions of 65 and 60 Å of CBH I and CBH II respective cores, would result in ~70 Å for the CBD, a value exceeding that of the cores which yet count more than 10 times as many residues. Inversely, assuming a dimension of 26 Å for the CBD (compatible with the 3D-structure [13]) and taking the measured dimensions of CBH I and CBH II cores (65 and 60 Å, respectively), linker lengths of 87 and 122 Å are derived for the two cellobiohydrolases. This would give 3.8 and 3.05 Å per residue for the linkers of CBH I and CBH II, respectively. There is nothing in the sequence of the linker of CBH I that would suggest a rise per residue larger than that of CBH II and/or a conformation more extended than that of a β-strand.

4. Conclusion

This paper shows the first application of dynamic light scattering to the study of a two-domain cellulase such as EG V of *H. insolens*. The hydrodynamic dimensions of the entire enzyme can be approximated as an ellipsoid with dimensions of 42 × 133.6 Å. This method should now enable the accurate determination of the hydrodynamic size of EG V mutants as to probe the possible flexibility between the constitutive domains. This flexibility (or stiffness) is likely to be found in many cell-wall degrading enzymes that have a catalytic domain carrying a cellulose-binding domain via long linkers rich in serine, threonine and proline.

Acknowledegments: The authors wish to thank A. Patkowski (Mainz) for stimulating discussions. This work was funded by a joined ADEME/ Novo-Nordisk fellowship to C.B. and the BIO2 CT 94 3018 research grant from the EC.

References

[1] van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R. and Pettersson, G. (1986) FEBS Lett. 204, 223–227.
[2] Gilkes, N.R., Warren, R.A.J., Miller Jr., R.C. and Kilburn, D.G. (1988) J. Biol. Chem. 263, 10401–10407.

[3] Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C. and Warren, R.A.J. (1991) Microbiol. Rev. 55, 303–315.
[4] Gilbert, H.J. and Hazlewood, G.P. (1993) J. Gen. Microbiol. 139, 187–194.
[5] Tomme, P., van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. and Claeyssens, M. (1988) Eur. J. Biochem. 170, 575–581.
[6] Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C. and Jones, T.A. (1990) Science 249, 380–386.
[7] Juy, M., Amit, A.G., Alzari, P.M., Poljak, R.J., Claeyssens, M., Béguin, P. and Aubert, J.-P. (1992) Nature 357, 89–91.
[8] Davies, G.J., Dodson, G.G., Hubbard, R.E., Tolley, S.P, Dauter, Z., Wilson, K.S., Hjort, C., Mikkelsen, J.M., Rasmussen, G. and Schülein, M. (1993) Nature 365, 362–364.
[9] Spezio, M., Wilson, D.B. and Karplus, P.A. (1993) Biochemistry 32, 9906–9916.
[10] Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K.C., Teeri, T.T. and Jones, T.A. (1994) Science 265, 524–528.
[11] White, A., Withers, S.G., Gilkes, N.R. and Rose, D.R. (1994) Biochemistry 33, 12546–12552.
[12] Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z., Wilson, K.S., Chauvaux, S., Béguin, P. and Alzari, P. (1995) Nature Struct. Biol. 2, 569–576.
[13] Kraulis, P.J., Clore, G.M., Nilges, M., Jones, T.A., Pettersson, G., Knowles, J.K.C. and Gronenborn, A.M. (1989) Biochemistry 28, 7241–7257.
[14] Xu, G.Y., Ong, E., Gilkes, N.R., Kilburn, D.G., Muhandiram, D.R., Harrisbrandts, M., Carver, J.P., Kay, L.E. and Harvey, T.S. (1995) Biochemistry 34, 6993–7009.
[15] Schmuck, M., Pilz, I., Hayn, M. and Esterbauer, H. (1986) Biotechnol. Lett. 8, 397–402.
[16] Abuja, P.M., Pilz, I., Claeyssens, M. and Tomme, P. (1988) Biochem. Biophys. Res. Commun. 156, 180–185.
[17] Abuja, P.M., Schmuck, M., Pilz, I., Tomme, P., Claeyssens, M. and Esterbauer, H. (1988) Eur. Biophys. J. 15, 339–342.
[18] Rasmussen, G., Mikkelsen, J.M., Schülein, M., Patkar, S.A., Hagen, F., Hjort, C.M. and Hastrup, S. (1991) World Patent Application WO 91/17243.
[19] Chu, B. (1991) Laser Light Scattering. Basic Principles and Practice, 2nd edn., Academic Press, San Diego.
[20] Provencher, S.W., Hendrix, J., De Maeyer, L. and Paulussen, N.J. (1978) J. Chem. Phys. 69, 4273–4276.
[21] Pecora, R. (1985) Dynamic Light Scattering, Plenum Press, New York.
[22] Koenig, S.H. (1975) Biopolymers 14, 2421–2423.