Dimension, shape and conformational flexibility of a two-domain fungal cellulase in solution probed by small angle X-ray scattering *

Véronique Receveur#, Mirjam Czjzek#, Martin Schülein ‡§, Pierre Panine¶, Bernard Henrissat#

#Architecture et Fonction des Macromolécules Biologiques, UMR 6098, CNRS and Universités d’Aix-Marseille I and II, 31 Chemin Joseph Aiguier, 13402 Marseille cedex 20, France.
‡ Novozymes, Smoermosevej 25, DK-2880 Bagsvaerd, Denmark
¶ European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble Cedex, France
§deceased

Corresponding author: Véronique Receveur.
Tel. +33 491 16 45 15;
Fax. +33 491 16 45 36; e-mail: receveur@afmb.cnrs-mrs.fr

Running Title : Dimension, shape and flexibility of a cellulase
Summary

Cellulase Cel45 from *Humicola insolens* has a modular structure with a catalytic module and a cellulose-binding module (CBM) separated by a 36 amino acid, glycosylated, linker peptide. The solution conformation of the entire two-domain Cel45 protein as well as the effect of the length and flexibility of the linker on the spatial arrangement of the constitutive modules were studied by small angle X-ray scattering combined with the known 3-D structure of the individual modules. The measured dimensions of the enzyme show that the linker exhibits an extended conformation leading to a maximum extension between the two centres of mass of each module corresponding to about 4 cellobiose units on a cellulose chain. The glycosylation of the linker is the key-factor defining its extended conformation, and a 5-proline stretch mutation on the linker was found to confer a higher rigidity to the enzyme. Our study shows that the respective positioning of the catalytic module and the CBM onto the insoluble substrate is most likely influenced by the linker structure and flexibility. Our results are consistent with a model where cellulases can move on the surface of cellulose with a caterpillar-like displacement with free energy restrictions.
Introduction

The enzymatic hydrolysis of cellulose, the most abundant biopolymer on earth, is catalyzed by cellulases, which have considerable industrial and ecological importance. Most cellulases have a modular structure with a catalytic module and a cellulose-binding module (CBM) usually joined by a glycosylated and presumably flexible linker peptide (1). Removal of the CBM results in a significantly reduced activity of the enzymes on crystalline cellulose probably because of a decreased binding capacity, but the activity on soluble cellulose oligomers is retained (2-4). Inversely, isolated CBMs retain most of the binding capacity on crystalline cellulose, but are devoid of catalytic action (3, 5). Enzymatic activity is sometimes affected when the interdomain linker has been shortened or completely deleted suggesting that it should be of sufficient length and/or flexibility to ensure an independent action of the two functional modules (6, 7).

The catalytic modules of cellulases have been classified into several distinct families of glycoside hydrolases on the basis of amino acid sequence similarities (8, 9) while, similarly, the CBMs form several families (10, 11). Structural data on individual cellulase modules have been accumulating over the last few years, with 3-D structures solved in nine families of catalytic modules and seven families of CBMs (12). There is no 3-D structure available for an intact cellulase whose catalytic module and CBM are separated by an interdomain linker longer than a dozen residues. The larger length, flexibility and heterogeneous glycosylation of fungal interdomain linkers are probably a major obstacle to protein crystallization. The only two-domain cellulase known at the 3-D structural level is that of the bacterial endoglucanase E4 (Cel9A) of Thermobifida fusca (13). However, in this particular protein, the interdomain linker is so short that the CBM is literally ‘fused’ onto the catalytic module.
So far, all the different experimental techniques used for structural elucidation failed to give precise information on the overall structure of full length, multidomain cellulases bearing a long, glycosylated and presumably flexible interdomain linker. At the end of the 80’s, Pilz and co-workers have reported pioneering studies using small angle X-ray scattering (SAXS), which provided the very first insight into the tertiary structure of two-domain cellulases from *Trichoderma reesei* (14, 15) and from *Cellulomonas fimi* (6, 16, 17). However, these early results were necessarily imperfect due to the lack of known 3-D structure for the isolated cellulase modules and due to the equipment available then. The data collected on conventional sources in the 80’s were undermined by relatively poor statistics, despite the huge amounts of protein used for the experiments (several hundred mg for one set of experiment). Besides, the high protein concentrations used may have distorted the signal (18, 19).

Since then, only the works of Boisset and co-workers have combined the structural information now available on isolated modules with dynamic light scattering experiments to establish a more precise estimate of the hydrodynamic dimensions of a full length cellulase (20). However, it is difficult to obtain precise structural information on proteins from dynamic light scattering data as only rough models based on geometrical shape can be used. Moreover the choice of a model requires to rely on preliminary assumptions. By contrast, Small Angle X-ray Scattering (SAXS) is a method of choice to determine the average shape of a protein because it enables to measure directly the dimensions of a macromolecule in solution and to calculate its envelope *ab initio*, without any starting hypothesis. Here, we report the first *ab initio* characterization of the two-domain structure of cellulase Cel45 from the fungus *Humicola insolens* and of several of its variants by SAXS in order to shed light on the shape, dimension and conformation of this modular protein. Cel45 has a catalytic module belonging to glycoside hydrolase family 45 carrying a CBM of family 1 via a 36 amino acid long linker
peptide (21). The three-dimensional structure of the catalytic module has been determined (22) and that of the CBM could be directly modeled from its 50% sequence identity to that of cellobiohydrolase I (Cel7A) from Trichoderma reesei whose 3-D structure was solved by NMR (23).
Experimental procedures

Sample Preparation

*H. insolens* Cel45 full-length and its isolated catalytic module (Cel45 core) were cloned and expressed in *Aspergillus oryzae* (24) and purified as already described (20). Site directed mutagenesis was done using the PCR method. The expression plasmids harbouring the mutated genes were transformed into *Aspergillus oryzae* using selection on acetamide by cotransformation using the AMDs selection as already described (25). A variant comprising only the catalytic module and the linker (variant Cel45ΔCBM) was obtained by introducing a stop codon on the Cel45 encoding gene at a position corresponding to amino acid 247. Another variant featuring a stretch of 5 consecutive proline residues in the interdomain linker (variant Cel45 PP) was obtained by changing the codons for Val-240 and Gln-241 to codons for prolines. Finally, a variant having a half-size linker (variant Cel45 ΔS219-T235 ) was obtained by deletion of the region coding from Ser-219 to Thr-235. Variants Cel45 ΔS219-T235 and Cel45 PP, which bear a CBM, were purified using Avicel affinity as described earlier (20). Variant Cel45ΔCBM was purified using ion-exchange chromatography as described for the catalytic module (Cel45 core) (20).

The lyophilised proteins were diluted in 50mM sodium phosphate buffer, pH 7.5, and then washed extensively through a microconcentrator equipped with a Filtron® polyvinylidene membrane in order to remove contaminating salts. Finally, each protein sample was filtered through a Rainin Microfilterfuge membrane (100,000 molecular weight cut-off) in order to eliminate non-specific aggregates from the solution, and the filtrate was used for the SAXS experiments. The protein concentration of each sample was checked by absorbance at 280 nm using extinction coefficients calculated from the amino acid
composition of each protein. Glycerol (1% and 10%, v/v), was added to the buffer as a radiation scavenger for SAXS experiments performed at LURE and ESRF respectively.

**SAXS experiments**

SAXS experiments were carried out on the D24 instrument installed on the bending magnet of the storage ring LURE-DCI (Orsay, France) and on beamline ID2 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). On D24, the wavelength $\lambda$ was 1.488 Å and the sample-to-detector distances were 1.32 m and 2.08 m. The vertical linear-sensitive position detector was shifted upwards with respect to the incident beam so that the scattered intensity for higher scattering angles could be measured. This setup gave access to scattering vectors $q$ ranging from 0.015 to 0.35 Å$^{-1}$ and 0.0009 to 0.22 Å$^{-1}$ respectively. The scattering vector is defined as $q=4\pi/\lambda \sin \theta$, where $2\theta$ is the scattering angle. Eight to sixteen frames of 200s were recorded depending on the protein concentration. The protein solution was continuously circulated through the evacuated quartz capillary so that no radiation-induced damage was observable. On ID2, the wavelength was 1.0 Å and the sample-to-detector distance was 3.0 m, leading to scattering vectors $q$ ranging from 0.01 to 0.022 Å$^{-1}$. The detector was a X-Ray Intensified optically coupled CCD camera, and 100 successive frames of 1 second with 5 seconds pause between each frame were recorded for each sample. A stopped-flow equipment (Bio-Logic, Claix, France) was used to mix the protein and the buffer solution at a chosen concentration, and a new solution was injected before each frame into a home-built observation cuvette made from a 1.5mm Lindemann-type quartz capillary. Thus, no protein solution was irradiated longer than one second. Each frame was then carefully inspected to check for possible bubble formation or radiation-induced aggregation. No such effect was observed and individual frames could then be averaged. Absolute calibration was made with a Lupolen sample.
A series of measurements at different protein concentrations ranging from 1 to 11 mg/ml were performed for each protein to check for interparticle interaction. Background scattering was measured before or after each protein sample using the corresponding buffer solution and then subtracted from the protein scattering patterns after proper normalisation and correction from detector response. All the experiments were conducted at 20°C.

**Scattering Data Analysis**

The values of radii of gyration (R_g) were derived from the Guinier approximation (26): 

\[ I(q) = I(0) \exp(-q^2 R_g^2/3) \]

where \( I(q) \) is the scattered intensity and \( I(0) \) is the forward scattered intensity. The radius of gyration and \( I(0) \) are inferred respectively from the slope and the intercept of the linear fit of \( \ln[I(q)] vs q^2 \) at low q values.

The distance distribution function \( P(r) \) indicates the distribution of the interatomic vectors of a molecule with a given length. This enables an alternative calculation of R_g using the entire scattering curve (R_g is estimated from the second moment of the \( P(r) \) function). The function also provides the maximum dimension \( D_{max} \) of the molecule, which is defined as the point where \( P(r) \) becomes zero. Distance distribution functions were calculated by the Fourier inversion of the scattering intensity \( I(q) \) using GNOM (27). A range of \( D_{max} \) was tested and the final choice of \( D_{max} \) was based on three criteria: \( P(r) \) should exhibit positive values, the R_g value from GNOM should agree with that from the Guinier analyses, and the \( P(r) \) curve should remain stable when increasing \( D_{max} \) beyond the estimated macromolecular length.

Calculations of the theoretical scattering functions based on crystal structures were performed using CRYSOL (28). They were fitted to the experimental data in the case of the isolated catalytic module by taking into account the excluded volume and the scattering by a hydration shell having a slightly different density than the bulk solvent. The structural model
for Cel45 core was PDB file with ID code 2eng. The coordinates for the CBM were calculated by homology modelling with SwissModel (29) using the coordinates of the CBM from \textit{T. reesei} cellobiohydrolase (PDB ID code: 1cbh) (23).

\textbf{3-D Modelling}

The low-resolution shapes of Cel45 variants and wild type were determined \textit{ab initio} from the scattering curves using the programs DAMMIN (30) and GASBOR (31). These two programs restore low-resolution shapes of protein and calculate a volume filled with densely packed spheres (dummy atoms or dummy residues respectively) fitting the experimental scattering curve by a simulated annealing minimisation procedure.

DAMMIN minimises the interfacial area between the molecule and the solvent by imposing compactness and connectivity constraints and can fit the scattering curve up to a resolution of 20\AA. It automatically subtracts a constant from the scattering profile in order to force it to follow the Porod Law (32) as the surface between the particle and the solvent should be well-defined. GASBOR calculates a 3-D model from the entire scattering pattern considering the protein as an assembly of dummy residues centred on the C\textalpha positions (spheres of 3.8\AA diameter) with a nearest-neighbour distribution constraint. This provides more reliable and higher resolution models than DAMMIN if the experimental data are of sufficient quality.

In our case, the cellulases are heavily O-glycosylated in the linker region. To take into account this glycosylation we have considered, for GASBOR, that a sugar residue is equivalent to \textasciitilde1.6 amino acid residue according to its electronic density and to its length. We have then calculated the total number of real and equivalent amino acid residues of the cellulase knowing its primary structure and the average number of glycosides contained in the
protein given by mass spectrometry measurements. We used this global number of residues for the protein in the input of the GASBOR program.

In every case, several independent fits were run with DAMMIN (spheres of diameters 6, 5 and 4Å) and GASBOR with no symmetry restriction and the stability of the solution was checked. The isolated modules were then positioned in the envelope with respect to each other using TURBO-FRODO (33).
Results

Intact two-domain Cel45 as well as its isolated catalytic module, and three different variants altered in the interdomain linker have been studied by SAXS in order to estimate the influence of the linker on the global conformation of this cellulase. In variant Cel45 PP, two amino acids of the wild-type linker were replaced by two prolines resulting in a sub-sequence of 5 consecutive proline residues. This short sequence, which adopts a single preferred conformation in solution known as a polyproline II helix (34), was devised in order to increase the rigidity of the linker. In variant Cel45 ΔS219-T235, the catalytic module and the CBM are joined together by a linker 17 amino acids shorter than the wild-type linker. Finally, in variant Cel45ΔCBM, the CBM has been removed resulting in a protein comprising only the catalytic module and the full length linker peptide.

Linker glycosylation

The molecular weight of Cel45 and its variants has been estimated by MALDI-TOF mass spectrometry (Martin Schülein, unpublished). Broad peaks were observed for all samples except for the isolated catalytic core due to the heterogeneous O-glycosylation of the linker. A comparison of the average molecular weight observed for each protein with that calculated from the amino acid composition gave an estimate of the overall glycosylation content of the linker: ~ 40 sugar units for Cel45, Cel45 PP and Cel45ΔCBM and ~16 sugar units for variant Cel45 ΔS219-T235. This approximates to an average of about 1.7 sugar per Ser or Thr residue in the linker. The obvious glycosylation heterogeneity is likely to be accompanied by an heterogeneity of conformations in solution.
**Guinier analysis**

The average size of Cel45 and its variants was estimated by the measure of their radii of gyration. Figure 1 shows the experimental SAXS profiles in Guinier plots for all studied proteins. The scattered intensities follow closely the Guinier law in the small-\( q \) region and no sign of aggregation can be observed. The Guinier approximation is often used in the \( q \)-range corresponding to \( q \cdot \text{Rg} \leq 1.3 \). Because the proteins studied here are probably not spherical and present a distribution of conformations, the Guinier approximation was therefore considered to be valid only up to \( q \cdot \text{Rg} \leq 1 \). No concentration effect was detected, except for Cel45 PP, for which \( \text{Rg} \) was extrapolated to infinite dilution. The radii of gyration of Cel45 and of its variants are reported in Table 1. The values of radii of gyration obtained for the proteins bearing a linker are much higher than what would be expected for a spherical protein with the same number of amino acids (35). This is particularly evident when one compares the radii of gyration of the isolated catalytic module (17.3 Å, 210 amino acids) with variant Cel45ΔCBM which consists of the catalytic module and the linker only. The latter exhibits a \( \text{Rg} \) of 30.0 Å although it is larger than the former by only 36 amino acids and about 40 sugar units. By contrast, the CBM (38 amino acids) does not contribute much to the average size of the protein since the intact Cel45 displays a \( \text{Rg} \) of 33.6 ± 0.4Å. Interestingly, the radius of gyration of variant Cel45 PP (\( \text{Rg}=35.5 \pm 0.5\)Å) is slightly but significantly larger than that of wild-type Cel45, indicating a substantial difference in their average size for the two proteins even though their primary structures differ only by two amino acids.

**Distance distribution function**

Distance distribution functions were deduced from the scattering intensities of Cel45 and its variants. The experimental \( P(r) \) curves and the theoretical \( P(r) \) of the CBM calculated from its atomic coordinates are presented in Figure 2. The distance distribution function of the isolated
catalytic module has a bell-shape, which is characteristic of a globular protein. The experimental profile corresponds exactly to the theoretical one calculated from the crystal structure coordinates with a maximum dimension of 45±5 Å.

The $P(r)$ profiles of full length Cel45 and of its variants display a biphasic pattern, clearly indicative of an elongated shape according also to the large maximum dimension obtained: 125 ± 5 Å for full length Cel45, 122 ± 4 Å for variant Cel45 PP, 100 ± 10 Å for variant Cel45 ∆CBM and 95 ± 10 Å for variant Cel45 ∆S219-T235. This is consistent with the high values of radii of gyration measured for these molecules with respect to their molecular mass. The first peak is representative of the short intramolecular distances, mainly within the catalytic module, as evident from the superimposition on the catalytic module $P(r)$ profile. The broadening to the right of the first peak indicates that there are larger distances within the parts of the cellulase other than the catalytic module or the CBM. This suggests that the linker region is rather extended. The second part of the curve characterises interdomain distances. This second broad peak corresponds to the distances between the catalytic module and the linker region containing the peptide chain and the carbohydrate residues, as confirmed by the distance distribution function of Cel45∆CBM. The last extension of the curve observed in wild-type Cel45 and in its CBM-containing variants corresponds to the distances between the CBM and the catalytic module and indicates that they sit at opposite ends of the molecule. The $D_{\text{max}}$ inferred from all these curves are reported in Table 1.

The $P(r)$ curves of Cel45 WT and of variant Cel45 ∆S219-T235 (figure 2b) have a similar pattern and differ only by the extent of the second peak. On the contrary, the wingspan of Cel45 WT and variant Cel45 PP are identical but the $P(r)$ profiles are not the same. These differences on the intermediate distances suggest either that the orientation of the two main modules with respect to each other is dissimilar in the two proteins (36), or that the two proteins display specific chain conformations in the hinge region.
Three-dimensional models for Cel45 and its variants

*Ab initio* 3-D envelopes of Cel45 and its variants have been determined using the program packages DAMMIN and GASBOR. DAMMIN is particularly appropriate for objects displaying a sharp interface with the solvent and which follow the Porod law. However, the surface of a cellulase is not very well defined because of the loose compactness of the linker peptide and of the glycoside side chains. Indeed this group of molecules probably does not follow the Porod law. Therefore, we also used GASBOR even though this program is firstly intended for higher resolution data. Repeated fits were computed independently with the two programs. In fact, due to the intrinsic limitations of the scattering technique on solution samples, *ab initio* 3D shape restoration from 1D scattering profiles can unavoidably lead to incorrect structures fitting the data even so. Thus, repetitions of the fit, even if it cannot totally prevent it, significantly decreases the risk to infer an erroneous structure.

The results obtained with both programs were superimposed using SUPCOMB (37). This showed that the calculated envelopes are nearly identical for each protein whatever the program used. The best models provided a fit to the experimental data with $\chi^2=2.3$ for Cel45 PP, 2.1 for Cel45 ΔS219-T235 and 0.97 for Cel45 WT. The fits and the resulting shapes are presented on figure 3.

The *ab initio* shapes reveal elongated particles, with a globular spherical part and a more or less long and thick protuberance. The prominence is rather cylindrical in the Cel45 ΔS219-T235 variant, whereas its cross section diameter varies in the Cel45 PP variant with tight and swollen regions in the middle part. This outgrowth is linear in Cel45 ΔS219-T235 variant, exhibits a bend in Cel45 PP variant, and several bends in different orientations in Cel45WT. These geometries are perfectly reproducible from one fit to another with GASBOR and DAMMIN for the two variants, while there are some minor differences in the bending of
the linker in the models obtained for Cel45 WT. The P(r) profiles of Cel45 core, Cel45ΔCBM and full length Cel45 allowed us to position the isolated modules in these envelopes at opposite ends of the elongated shape. The central part of the envelope therefore represents the glycosylated linker peptide.
Discussion

Dimensions of the whole cellulase

The SAXS experiments on cellulase Cel45 and variants reported here enabled us to determine the global structural characteristics of a full length cellulase and to obtain the first \textit{ab initio} shape of a two-domain cellulase. Our results considerably improve the first models of cellulases that Pilz and co-workers derived at a time when the composition and structure of individual modules were not known (14-16). Here we show that the shape of cellulase Cel45 is indeed similar to a tadpole. However the dimensions of the linker peptide inferred from our experiments are considerably more accurate and consistent with the number of residues in the linker. As a matter of fact the maximum dimensions obtained for CBHI and CBHII from \textit{T. reesei} were probably overestimated and lead to unrealistic values for the dimension of the linker: $\sim105\text{Å}$ and $\sim125\text{Å}$ respectively for the linkers of CBHI (28 residues) and CBHII (41 residues). By contrast, our results are consistent with the work of Boisset and co-workers (20) on \textit{H. insolens} Cel45 using dynamic light scattering (estimated $D_{\text{max}}$ of the whole protein of 133Å).

Flexibility of the cellulase

It is important to remind that a scattering experiment measures the average of all species in the solution spot irradiated by the beam during the frame time. Therefore a scattering curve may correspond to the mixture of different protein conformations if the structure exhibits some flexibility. It is well known that the radius of gyration is the root-mean squared distance of all atoms to the centre of the scattering volume of the particles in solution, weighted by their scattering length. In other words, Rg describes the average conformation adopted by all the protein molecules in solution. In contrast, $D_{\text{max}}$, obtained from the distance distribution function, corresponds to the maximum distance existing between two atoms within a
molecule amongst all conformations present in solution. It is therefore very informative to confront these two parameters for the different variants.

The values of the radii of gyration and of the \(D_{\text{max}}\) of Cel45 wild type both indicate that the dimension of the linker is very large with respect to its mass. \(D_{\text{max}}\) and \(R_g\) considerably increase from Cel45 core to variant Cel45\(\Delta\)CBM. The linker, therefore, is not compact but has an oblong shape protruding to the exterior and is not wrapped around the catalytic module. Furthermore, on the basis of the \(P(r)\) profiles, one can conclude that the presence of the CBM does not seem to alter the global conformation of the linker of the cellulase in solution, since the observed increase of \(D_{\text{max}}\) corresponds exactly to the dimension of the CBM calculated from its modeled structure.

When the linker is shortened by 17 amino acids (variant Cel45 \(\Delta S219-T235\) ), the \(R_g\) and \(D_{\text{max}}\) values decrease considerably, corresponding to a loss of 30 Å in length. The resulting value of about two residues per Å indicates that the linker can adopt a rather extended conformation instead of a compact, globular structure stabilised by short distance interactions. This suggests that the linker region is rather flexible. The diffusion curves are representative of a mixture of all possible conformations present in solution.

Our results demonstrate that the linker occupies a large volume with respect to its relative small number of amino-acids. Whereas the catalytic module and the CBM have an ordered globular structure, the linker is not compact at all and must exhibit some internal flexibility. Its low atomic density indicates that short distance interactions cannot induce the cohesion of the peptide. However, hydrogen bonds (through water molecules or not) may generate some local structured clusters. Furthermore the important transversal steric hindrance due to the glycosylation of the linker probably restricts the conformational space available. Because of the heterogenous nature of O-glycosylation (in sugar chain length and position), different steric constraints must apply within each protein molecule in the solution, allowing
or forbidding some torsion angles on the peptide chain and therefore affecting some of the conformations that the linker can adopt.

Another remarkable feature is that the Rg values of wild type Cel45 and of variant Cel45 PP are different even though their $D_{\text{max}}$ are identical and they have the same number of amino acids in the linker region. The maximum extension of the linker is the same whereas the average size is smaller in the case of wild-type Cel45. Furthermore, the P(r) profile indicates that there is a lower number of intermediate distances for variant Cel45 PP (Figure 2). The ab initio calculated shapes are also quite different in the linker region. In Cel45 WT and variant Cel45 ΔS219-T235, the thickness of the linker is rather constant, indicating a homogenous distribution of the glycosylation along the linker and of the possible conformations it can adopt. On the contrary, we can observe a narrowing of the linker close to the catalytic module in variant Cel45 PP with a broadening at higher distances. As no mutation affected the glycosylation sites in Cel45 PP compared to Cel45 WT, this suggests that the flexibility of the linker of Cel45 PP is not homogeneous along the linker. Since Cel45 PP adopts average conformations more extended than the wild type, but still with an identical maximum possible extension, we conclude that Cel45 WT adopts compact conformations more easily than Cel45 PP and that the five-proline stretch introduced in the linker indeed confers a higher global rigidity to the protein.

**Linker and CBM functions**

The distance at which the catalytic module can operate from the adsorption site of the CBM on cellulose is an essential parameter to understand the synergistic function of the isolated modules of cellulases. Cellulose consists of glucose residues connected by β-1,4 glycosidic bonds. In crystalline cellulose, each residue is rotated by 180° with respect to the previous one, resulting in a cellobiose repeat unit 10.4 Å long. For Cel45 we have obtained a $D_{\text{max}}$ of
125 Å which results in a maximum distance between the centres of the constitutive modules of 85 ± 10 Å (Fig. 10). From this, one can conclude that from a fixed binding site of the CBM onto a cellulose chain, Cel45 has a maximum operation range of about 40 Å, e.g. may hydrolyse only a maximum of four glycosidic bonds of similar orientation on a single chain. This limited range would be a serious impediment as the bonds cleavable by one catalytic domain would rapidly be exhausted. It has been shown that CBMs are not fixed on the cellulose surface, but instead that they are quite mobile and able to diffuse on the substrate surface (38). Our results therefore reinforce this view where the mobility of CBMs on the cellulose surface enables an efficient hydrolysis by the catalytic module (38, 39). Srisodsuk et al. have shown that a sufficient spatial separation between the two modules is required for optimal activity in the case of T. reesei CBHI (7). This means that the linker certainly optimises the geometry between the two modules with respect to the substrate. Thus, our results are consistent with a model where cellulases can move on cellulose following a caterpillar-like motion. While the CBM remains attached to a specific site, the catalytic module can hydrolyse several glycosidic bonds within the 40 Å range allowed by the flexibility of the linker. When the linker eventually becomes too tightly compressed (or extended), the free energy of the cellulase might be released via the translation of the CBM along the cellulose surface. Because the free energy is likely to increase both upon extension or compression of the linker, the crawling motion does not depend on the processivity direction of the enzyme (from the reducing end to the other or vice versa) nor on the position of the CBM at the N- or C-terminal end of the cellulase. Such a model is consistent with the view of Linder & Teeri who already suggested that the two modules should act in concert on the cellulose surface during catalysis and that relatively long linker regions with some flexibility are needed to express full cellulolytic activity (40).
The present paper builds on the excellent complementarity of the SAXS technique with protein crystallography and NMR for the study of the conformation of modular proteins. The work reported here establishes the first precise model of a two-domain cellulase in solution combining small angle X-ray scattering data with the now available 3-D structures of each individual globular module. Our results thus provide the first information on the structural properties of the interdomain peptide probed with a spectroscopic technique and show that the linker is flexible and extended but still exhibits some structural constraints. The amino acid composition of the linker plays an important role as shown by our results on variant Cel45 PP. Our SAXS data show that the insertion of a polyproline sequence actually leads to a higher rigidity of the linker. Another parameter controlling the flexibility of the linker is certainly the amount of glycosylation. A high glycosylation not only may be necessary to protect the linker from proteolytic cleavage and to stabilise the whole protein, but it also appears to enable the linker to attain an extended structure. It is therefore likely that non-glycosylated or underglycosylated linkers will be more flexible than those heavily glycosylated. Further studies of the role of the glycosylation on the structure and flexibility of the linker in cellulases are underway in the laboratory. Finally, the present analysis of the conformational behaviour of a two-domain cellulase shed new light on the mechanism of action of cellulases on the substrate surface and how their catalytic modules and CBD co-operate.
Acknowledgements

We acknowledge the storage ring LURE-DCI (Orsay, France) for provision of synchrotron radiation facilities and we would like to thank Patrice Vachette and Javier Perez for assistance in using beamline D24 and for valuable discussions. We also thank Alain Roussel for his help with TURBO-FRODO.
References

1. Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. (1991) *Microbiol. Rev.* **55**, 303-315.

2. Gilkes, N. R., Warren, R. A., Miller, R. C., Jr. & Kilburn, D. G. (1988) *J. Biol. Chem.* **263**, 10401-10407.

3. van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R. & Pettersson, G. (1986) *FEBS Lett.* **204**, 223-227.

4. 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.

5. Gilkes, N. R., Jervis, E., Henrissat, B., Tekant, B., Miller, R. C., Jr., Warren, R. A. & Kilburn, D. G. (1992) *J. Biol. Chem.* **267**, 6743-6749.

6. Shen, H., Schmuck, M., Pilz, I., Gilkes, N. R., Kilburn, D. G., Miller, R. C. & Warren, R. A. J. (1991) *J. Biol. Chem.* **266**, 11335-11340.

7. Srisodsuk, M., Reinikainen, T., Penttila, M. & Teeri, T. T. (1993) *J. Biol. Chem.* **268**, 20756-20761.

8. Henrissat, B. & Bairoch, A. (1993) *Biochem. J.* **293**, 781-788.

9. Henrissat, B. & Bairoch, A. (1996) *Biochem. J.* **316**, 695-696.

10. Tomme, P., Warren, R. A. J., Miller, R. C. J., Kilburn, D. G. & Gilkes, N. R. (1995) in *Enzymatic Degradation of Insoluble Polysaccharides*, eds. Saddler, J. N. & Penner, M. (American Chemical Society, Washington), Vol. 618, pp. 142-163.

11. Boraston, A. B., McLean, B. W., Kormos, J. M., Alam, M., Gilkes, N. R., Haynes, C. A., Tomme, P., Kilburn, D. G. & Warren, R. A. J. (1999) in *Recent Advances in Carbohydrate Bioengineering*, eds. Gilbert, H. J., Davies, G. J., Henrissat, B. & Svensson, B. (The Royal Society of Chemistry, Cambridge), pp. 202-211.
12. Bourne, Y. & Henrissat, B. (2001) *Curr. Opin. Struct. Biol.* **11**, 593-600.

13. Sakon, J., Irwin, D., Wilson, D. B. & Karplus, P. A. (1997) *Nat. Struct. Biol.* **4**, 810-818.

14. Abuja, P. M., Schmuck, M., Pilz, I., Tomme, P., Claeyssens, M. & Esterbauer, H. (1988) *Eur. Biophys. J.* **15**, 339-342.

15. Abuja, P. M., Pilz, I., Claeyssens, M. & Tomme, P. (1988) *Biochem. Biophys. Res. Commun.* **156**, 180-185.

16. Pilz, I., Schwarz, E., Kilburn, D. G., Miller, R. C., Warren, R. A. J. & Gilkes, N. R. (1990) *Biochem. J.* **271**, 277-280.

17. Meinke, A., Schmuck, M., Gilkes, N. R., Kilburn, D. G., Miller, R. C. & Warren, R. A. J. (1992) *Glycobiology* **2**, 321-326.

18. Zimmerman, S. B. & Minton, A. P. (1993) *Annu. Rev. Biophys. Biomol. Struct.* **22**, 27-65.

19. Receveur, V., Durand, D., Desmadril, M. & Calmettes, P. (1998) *FEBS Lett.* **426**, 57-61.

20. Boisset, C., Borsali, R., Schülein, M. & Henrissat, B. (1995) *FEBS Lett.* **376**, 49-52.

21. Schülein, M., Tikhomirov, D. & Schou, C. (1993) in *Trichoderma reesei cellulases and other hydrolases*, eds. Suominen, P. & Reinikainen, T. (Foundation for Biotechnical and Industrial Fermentation Research, Espoo), Vol. 8, pp. 109-116.

22. Davies, G. J., Dodson, G. G., Hubbard, R. E., Tolley, S. P., Dauter, Z., Wilson, K. S., Hjort, C., Mikkelsen, J. M., Rasmussen, G. & Schülein, M. (1993) *Nature* **365**, 362-364.

23. Kraulis, P. M., Clore, M. G., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J. & Gronenborn, A. M. (1989) *Biochemistry* **28**, 7241-7257.
24. Rasmussen, G., Mikkelsen, J. M., Schülein, M., Patkar, S. A., Hagen, F., Hjort, C. M. & Hastrup, S. (1991) in World Patent WO 91 17,243.
25. Christensen, O. T., Woldike, H., Boel, E., Mortensen, S. B., Hjortshoej, K., Thim, L. & Hansen, M. (1988) Bio/Technology 6, 1419-1422.
26. Guinier, A. & Fournet, F. (1955) Small angle scattering of X-rays (Wiley Interscience, New York).
27. Svergun, D. (1992) J. Appl. Cryst. 25, 495-503.
28. Svergun, D., Baraberato, C. & Koch, M. H. (1995) J. Appl. Cryst. 28, 768-773.
29. Peitsch, M. C. (1996) Biochem. Soc. Trans. 24, 274-279.
30. Svergun, D. I. (1999) Biophys. J. 76, 2879-2886.
31. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. (2001) Biophys. J. 80, 2946-2953.
32. Porod, G. (1982) in Small Angle X-ray Scattering, eds. Glatter, O. & Kratky, O. (Academic Press, London), pp. 17-51.
33. Roussel, A. & Cambillau, C. (1989) in Silicon Graphics Geometry Partners, ed. Committee, S. G. (Silicon Graphics, Mountain View, California), pp. 77-78.
34. Kelly, M. A., Chellgren, B. W., Rucker, A. L., Troutman, J. M., Fried, M. G., Miller, A. F. & Creamer, T. P. (2001) Biochemistry 40, 14376-14383.
35. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A. & Smith, L. J. (1999) Biochemistry 38, 16424-16431.
36. Glatter, O. (1982) in Small Angle X-ray Scattering, eds. Glatter, O. & Kratky, O. (Academic Press, London), pp. 167-196.
37. Kozin, M. B. & Svergun, D. I. (2000) J. Appl. Cryst. 34, 33-41.
38. Jervis, E. J., Haynes, C. A. & Kilburn, D. G. (1997) J. Biol. Chem. 272, 24016-24023.
39. Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P. & Gilbert, H. J. (1998) *Biochem. J.* **331**, 775-781.

40. Linder, M. & Teeri, T. T. (1997) *J. Biotechnol.* **57**, 15-28.
Footnotes

* This work was funded by contract BIO4-97-2303 of the European Commission and supported by the Centre National de la Recherche Scientifique.

1 Abbreviations used: CBM, cellulose-binding module; SAXS, small angle x-ray scattering; Cel45, *Humicola insolens* endoglucanase Cel45; R_g, radius of gyration; D_max, maximum dimension.
Figure legends

Figure 1.
Guinier plot of the scattered intensities of *Humicola insolens* Cel45 wild type and variants. From top to bottom: stars, Cel45 WT; filled triangles, Cel45 PP; open squares, Cel45 with no CBM; filled circles, Cel45 ΔS219-T235. Data are offset on the vertical axis for better visibility.

*(figure 1 should be at the beginning of the Results section, close to the ‘Guinier analysis’ sub-section)*

Figure 2.
Distance distribution function of *Humicola insolens* Cel45 WT and variants. (a) solid line, Cel45 WT; dashed line, Cel45 core; dotted line, Cel45 ΔCBM; Thick line, CBM (calculated from atomic coordinates). (b) solid line, Cel45 WT; dashed line, Cel45 PP; thin line, Cel45 ΔS219-T235.

*(figure 2 should be in the Results section, close to the ‘Distance distribution function’ sub-section)*

Figure 3.
Top: Calculated shapes of Cel45 WT (a), Cel45 PP (b) and Cel45 ΔS219-T235 (c); bottom: corresponding fits obtained with GASBOR on the scattering curves. The shapes are represented by their Connolly surfaces.

*(figure 3 should be at the end of the Results section)*
Table 1. Size and dimensions of Cel45 WT and variants

| Protein          | Mw (kDa) | Rg (Å)   | Dmax (Å) |
|------------------|----------|----------|----------|
| Cel45 WT         | 36       | 33.6±0.4 | 125 ± 5  |
| Cel45 core       | 22.9     | 17.3 ±0.3| 45 ± 5   |
| Cel45 ΔCBM       | 32       | 30.0 ±0.4| 100 ± 10 |
| Cel45 ΔS219-T235 | 31       | 23.1 ±0.5| 95 ± 10  |
| Cel45 PP         | 37       | 35.5 ±0.6| 122 ± 4  |
| Cel45 CBM        | 4^a      | 9.2^b    | 31^b     |

^a calculated from the sequence
^b calculated from the 3-D model

(Table 1 should be located in the results section, after figure 1 and before figure 2)
Figure 1
Figure 2
Figure 3
Dimension, shape and conformational flexibility of a two domain fungal cellulase in solution probed by small angle solution X-ray scattering
Véronique Receveur, Mirjam Czjzek, Martin Schulein, Pierre Panine and Bernard Henrissat

J. Biol. Chem. published online August 16, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M205404200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts