

Structural Insights into the Mechanism of Formation of Cellulosomes Probed by Small Angle X-ray Scattering*

Received for publication, August 5, 2004, and in revised form, October 19, 2004
Published, JBC Papers in Press, October 23, 2004, DOI 10.1074/jbc.M408979200

Michal Hammel‡§, Henri-Pierre Fierobe¶, Mirjam Czjzek‡, Stéphanie Finet||, Véronique Receveur-Bréchet‡**

From the ‡Architecture et Fonction des Macromolécules Biologiques, UMR 6098, CNRS and Universités d'Aix-Marseille I and II, 31 Chemin Joseph Aiguier, F-13402 Marseille cedex 20, France, ¶Bioénergétique et Ingénierie des Protéines UPR9036, CNRS, 31 Chemin Joseph Aiguier, F-13402 Marseille cedex 20, France, §Institute of Biophysics and X-ray Structure Research of the Austrian Academy of Sciences, Schmiedlstrasse 6, A-8042 Graz, Austria, and ||ESRF, 6 rue Jules Horowitz, BP 220, F-38043 Grenoble cedex 9, France

Exploring the mechanism by which the multiprotein complexes of cellulolytic organisms, the cellulosomes, attain their exceptional synergy is a challenge for biologists. We have studied the solution structures of the *Clostridium cellulolyticum* cellulosomal enzyme Cel48F in the free and complexed states with cohesins from *Clostridium thermocellum* and *Clostridium cellulolyticum* by small angle x-ray scattering in order to investigate the conformational events likely to occur upon complexation. The solution structure of the free cellulase indicates that the dockerin module is folded, whereas the linker connecting the catalytic module to the dockerin is extended and flexible. Remarkably, the docking of the different cohesins onto Cel48F leads to a pleating of the linker. The global structure determined here allowed modeling of the atomic structure of the *C. cellulolyticum* dockerin-cohesin interface, highlighting the local differences between both organisms responsible for the species specificity.

Cellulosomes are large (0.7–2 MDa) multiproteic complexes that efficiently degrade crystalline cellulose and related plant cell wall polysaccharides (1). They are produced by anaerobic cellulolytic bacteria, such as *Clostridium cellulolyticum* and *Clostridium thermocellum*, which are to date the best characterized. The cellulosomes enable these micro-organisms to grow on cellulose and to play a crucial role in the anaerobic recycling of carbon. The organization of these cellulolytic complexes has been investigated over the past decade and revealed very unusual and original features compared with the well known multiproteic complexes of similar mass, such as proteasomes (2) or ribosomes (3). Typically, clostridial cellulosomes are composed of a number of cellulases, or related enzymes, tenaciously bound to a scaffolding protein devoid of enzymatic activity (1). The latter is composed of a powerful cellulose binding module (CBM)¹ which anchors the whole complex at

the surface of the crystalline substrate, and generally of 3–11 “cohesin” modules that bind noncovalently to the catalytic subunits via a complementary module named “dockerin.” The cohesin/dockerin interaction is of high affinity ($K_a \geq 10^9 \text{ M}^{-1}$) and calcium-dependent (4–6). Interestingly, several reports have shown that at least within *C. cellulolyticum* and *C. thermocellum*, the cohesins are not specialized (*i.e.* any enzyme dockerin can bind to any cohesin of the scaffoldin with similar affinity) (6, 7). Thus, conversely to ribosomes or proteasomes, where each protein occupies a specific location in the complex, the cellulosomal enzymes are randomly incorporated into the multiprotein complex (8). For instance more than 20 different enzymes participate in the cellulosomes produced by *C. cellulolyticum* (9) and *C. thermocellum* (1), whereas their scaffoldins only contain 8 and 9 cohesins, respectively (7, 10). Thus, these bacteria concurrently secrete a whole population of cellulosomes with various enzyme compositions, stoichiometries, and arrangements (8, 9).

Nevertheless, the main paradox regarding the cellulosomes resides in their drastic enhanced catalytic efficiency, since the individual cellulosomal cellulases characterized to date display very low specific activity on crystalline cellulose. Therefore, exploring the mechanism of this synergy is a challenge for biologists. Recently, some evidence was obtained by exploiting the species specificity of the cohesin/dockerin interaction between *C. thermocellum* and *C. cellulolyticum* (11). Synthetic minicellulosomes were built containing two cellulases appended with either a *C. thermocellum* or a *C. cellulolyticum* dockerin bound in specific locations onto a hybrid miniscaffoldin, containing one cohesin from each species (5, 12). The incorporation of most tested enzyme pairs in these well defined minicellulosomes resulted in significant increase in activity toward model crystalline cellulose, such as Avicel. Analysis showed that the complexation of most cellulase pairs triggered synergies (up to 3.5-fold) between the two enzymes, whereas in the free state no or very low synergy was detected (12). This observation suggests that the binding to the cognate cohesin(s) induces a structural arrangement of the enzyme(s), leading to optimal activity and cooperation with the other catalytic subunits. The structural basis of this arrangement, leading to the synergistic properties observed upon binding to the cellulosome, is yet to be discovered.

The structures of a number of isolated cellulosomal modules or enzymes have been solved including the CBM (13, 14) and cohesin modules (15, 16) of the scaffoldins, produced by *lum*; Fcoeh and Ftcoeh, Fc and Ft, respectively, in complex with their cognate cohesin; NMA, normal mode analysis.

* This work was supported by the Centre National de la Recherche Scientifique, the Conseil Général des Bouches du Rhône, the Région Provence-Alpes-Côte d'Azur, and the Fonds zur Förderung der Wissenschaftlichen Forschung Grant J2220-N03 (to M. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** To whom correspondence should be addressed. Tel.: 33-4-91-16-45-15; Fax: 33-4-91-16-45-36; E-mail: receveur@afmb.cnrs-mrs.fr.

¹ The abbreviations used are: CBM, cellulose binding module; SAXS, small angle x-ray scattering; coheCt, cohesin 2 from *C. thermocellum*; coheCc, cohesin 1 from *C. cellulolyticum*; Fc, native full-length Cel48F; Ft, Cel48F variant with dockerin domain of Cel48S from *C. thermocel-*

C. thermocellum and *C. cellulolyticum*. Recently, the crystal structure of a *C. thermocellum* dockerin-cohesin complex has been established (17), providing precious information on the residues interacting in both modules. Nevertheless, to gain further insight into conformational events, likely to occur during the complexation of a cellulosomal enzyme with its cognate cohesin, structural studies have to be performed on an entire enzyme bound to the corresponding cohesin. So far, only truncated forms of cellulosomal enzymes lacking their dockerin domain have been successfully crystallized, and to our knowledge, crystals of entire cellulosomal enzymes bound to the corresponding cohesin have not been obtained yet, despite numerous attempts. Most probably, internal flexibility and conformational heterogeneity prevent the entire free cellulases or entire cellulases in the complexed state from crystallizing.

In the present study, a different approach, using small angle x-ray scattering (SAXS), has been used to investigate the mechanisms of formation of these complexes by analyzing the structure of an entire cellulosomal enzyme in free and complexed states. This technique is indeed a fundamental tool for the study of biological molecules in solution. The originality of this method is that it can provide structural information on molecules exhibiting some intrinsic disorder, flexibility, or heterogeneity, which usually constitute a major obstacle for the other classical structural methods (for a review, see Ref. 18). The probable flexibility of cellulosomal enzymes therefore incited us to resort to SAXS to elucidate their structural properties. The major cellulosomal cellulase Cel48F from *C. cellulolyticum* was chosen for this study, since the crystal structure of the catalytic module is available (19). We have examined different variants of the cellulase by SAXS: a truncated form lacking the dockerin domain, the wild-type entire form containing the native *C. cellulolyticum* dockerin at the C terminus, and an engineered cellulase appended with a *C. thermocellum* dockerin. The two types of Cel48F appended with either a *C. cellulolyticum* or a *C. thermocellum* dockerin were also studied in complex with their respective cognate cohesins. These constructs correspond to those used in minicellulosomes (5, 12). The three-dimensional rearrangements of the domains of the cellulosomal enzymes in free and complexed states were investigated according to the low resolution models restored from SAXS data, and the dynamic properties of these enzymes were explored, using normal mode analysis. Finally, a model for the three-dimensional atomic structure of the dockerin-cohesin complex of *C. cellulolyticum* was constructed based on our results. The physiological implication of these outcomes for the mechanisms of cellulosome formation and of their elevated synergy is then discussed.

EXPERIMENTAL PROCEDURES

Production and Purification of Recombinant Components

The construction of the plasmids pQE-Coh2 encoding cohesin 2 from *C. thermocellum* (coheCt), pET-coh1A encoding cohesin 1 from *C. cellulolyticum* (coheCc), pETFc encoding the native full-length Cel48F (Fc) from *C. cellulolyticum*, and pETFf encoding a modified Cel48F (Ft), in which the native dockerin domain is replaced by the dockerin domain from Cel48S of *C. thermocellum*, have already been described (5, 12). Production and two-step purification using nickel-nitrilotriacetic acid resin (Qiagen, Venlo, The Netherlands), followed by chromatography on Q-Sepharose fast flow (Amersham Biosciences), was performed essentially as formerly published (5, 12). In the case of the catalytic module of Cel48F, this was generated and purified from Fc; after the first chromatography on nickel-nitrilotriacetic acid resin, the protein sample was concentrated by ultrafiltration to 10 mg/ml and incubated at room temperature to promote the cleavage that occurs in the linker between the catalytic and the dockerin domains. Samples were periodically analyzed by SDS-PAGE to estimate the yield of the truncated form of the enzyme, and after 8 days of incubation, the proportion of truncated form was ~80%. The sample was loaded again on nickel-nitrilotriacetic

acid resin to remove the traces of the entire form and the dockerin domain. The purification of the catalytic module was finally achieved as described above (4, 5). The concentration of the purified proteins was estimated by quantitative amino acid analysis on a Beckman 6300 system (Fullerton, CA) using ninhydrin detection.

Preparation of Protein Samples for SAXS

All protein samples were prepared by dilution of the protein solutions to a final concentration of 10 mg/ml in 9.2 mM Tris-HCl, 1.84 mM CaCl₂, pH 8.0, 10% (v/v) glycerol. The complexes Fcohc and Ftcohc were prepared by mixing stoichiometric amounts of Fc and coheCc and of Ft and coheCt in the same buffer and at a final protein concentration of 10 mg/ml. Total complexation was checked by nondenaturing polyacrylamide gel electrophoresis on a Phast system apparatus (Phast gel gradient 4–15%, Amersham Biosciences). Immediately before the measurements, the protein samples were filtered through a Millex Microfilter membrane (pore size 0.22 μm) to eliminate existing large aggregates.

SAXS Experiments

SAXS measurements of isolated Cel48F catalytic module were carried out on a Kratky compact camera (HECUS Graz x-ray Systems, Austria) according to the procedure described by Hammel *et al.* (20).

Synchrotron SAXS experiments were performed at the European Synchrotron Radiation Facility (Grenoble, France) on beamline ID02. The wavelength λ was 1.0 Å. The sample-to-detector distances were set at 4.0 and 1.0 m, resulting in scattering vectors q ranging from 0.015 to 0.15 Å⁻¹ and 0.03 to 0.46 Å⁻¹, respectively. The scattering vector is defined as $q = 4\pi/\lambda \sin \theta$, where 2θ is the scattering angle. The detector was an x-ray-intensified optically coupled CCD camera, and 50 successive frames of 0.5-s exposure time (5-s interval between each frame) were recorded for each sample. To avoid radiation-induced protein damage, the enzyme solution was circulated through a quartz capillary. Each frame was carefully checked for possible bubble formation or radiation-induced aggregation. If such effects were not observed, the individual frames were averaged. Absolute calibration was made with a Lupolen sample. Background scattering was quantified before or after measurement of the protein sample and then subtracted from the protein patterns. All experiments were performed at 20 °C. The data acquired at both sample-to-detector distances of 4 and 1 m were merged for the calculations using the entire scattering spectrum.

Data Evaluation

The experimental SAXS data for all samples were linear in a Guinier plot in the low q region, indicating that the proteins do not undergo any aggregation. The radius of gyration, R_G , was derived by the Guinier approximation $I(q) = I(0) \exp(-q^2 R_G^2/3)$ up to $qR_G < 1.0$. The radii of gyration, R_G , calculated for different protein concentrations display no concentration dependence and indicate the absence of interaction in solution. The programs GNOM (21) and GIFT (22) were used to compute the pair-distance distribution functions $P(r)$. This approach gives the maximum dimension of the macromolecule D_{max} and offers an alternative calculation of R_G , which is based on the entire scattering curve. The program CRY SOL (23) was used to calculate SAXS profiles of the atomic structures and to compare them with the experimental scattering profile. The goodness of fit for all atomic models, as well as the low resolution models, with the experimental data was determined using the discrepancy χ , defined according to Konarev *et al.* (24).

Ab Initio Modeling

Modeling of the Overall Shapes—The overall shapes of the entire assemblies were restored from the experimental data by two independent programs: DAMMIN (25) and GASBOR (26). The scattering profiles were used up to $q_{max} = 0.46$ Å⁻¹ for the fit. The DAMMIN method calculates the scattering intensities of a particle built up from a finite number of dummy beads and minimizes the surface between the particle and the solvent. GASBOR searches a chain-compatible spatial distribution of an exact number of dummy residues, which corresponds to the C α atoms of protein amino acids. Five low resolution models obtained from different runs were averaged using the program DAMAVER (27) to construct the average model representing the general structural features of all the reconstruction.

Modeling of the Missing Domains and Loops—The program package CREDO (28) was used for adding missing domains or loops by fixing a known structure and building the unknown regions to fit the experimental scattering data obtained from the entire particle. This program was applied to all of our data to restore the low resolution models of the

TABLE I
 Experimental and model SAXS parameters

Construct	M_r^a	R_G^b	$R_G^{P(r)b}$	$R_G^{\text{Atomic}b}$	D_{max}	$\chi^{(\text{Over})c}$	$\chi^{(\text{CREDO})c}$	$\chi^{(\text{Sol})c}$
		Å	Å	Å	Å		single / average	
Cel48F catalytic module	70.8	24.8 ± 0.3	24.7 ± 0.1	24.7	74 ± 2			1.1
Fc	78.6	34.9 ± 0.8	33.4 ± 2.4	28.2	142 ± 6	1.0	2.2/1.8	2.4
Ft	80.5	36.0 ± 0.9	36.4 ± 1.5	29.2	147 ± 4	1.0	2.0/1.7	2.0
Fccoh	96.2	34.8 ± 0.6	35.2 ± 0.5	35.2	122 ± 3	0.7	0.8/0.6	1.6
Ftcoh	98.5	37.3 ± 0.9	37.9 ± 1.7	35.8	128 ± 4	0.7	1.5/1.3	1.7

^a Predicted from the sequence.

^b R_G , $R_G^{P(r)}$, and R_G^{Atomic} , radius of gyration given by the Guinier approximation, derived from the distance distribution function and calculated for the final atomic model using the program CRYSOLO, respectively.

^c $\chi^{(\text{Over})}$, $\chi^{(\text{CREDO})}$, and $\chi^{(\text{Sol})}$, discrepancies between the experimental SAXS profile and the fits calculated for the overall shapes-models, the fits calculated for the models restored by the program CREDO (the values given are the single fit and the average fits calculated by the program OLIGOMER), and the calculated scattering curve of the final solution structure, respectively.

dockerin and the dockerin-cohesin complex and to fit the experimental data obtained for the entire assemblies, in which the crystal structure of the catalytic module was used as the fixed starting point.

Construction of Solution Structures

Free Entire Enzymes Fc and Ft—To construct the solution structure of Fc and Ft, we used the crystal structure of the catalytic module of Cel48F attached to the low resolution model of the linker/dockerin region as modeled by the program CREDO. The atomic structure of the dockerin domain was superimposed with the CREDO model. The atomic coordinates of the missing linker between the dockerin and the catalytic module of Cel48F were modeled with the program TURBO starting from the C_α trace of the CREDO model as a template for the complete final model. Finally, to attest for the reliability of our modeling approaches, the theoretical scattering profiles obtained for the final atomic models were compared with the experimental SAXS data.

Overall Structure of the Entire Complexes Fccoh and Ftcoh—The atomic structures of the individual domains were positioned into the average overall shape of Fccoh and Ftcoh obtained by GASBOR using the program SUPCOMB (29). This arrangement was further refined and controlled using the program MASSHA (30). The program package CREDO was used to model the missing fragments corresponding to the linker between the catalytic module and the dockerin, and the loops in the cohesin, absent from the crystal structure, as well as the His tag fragments. Constructions of the final atomic models and the evaluation of these constructions were done by the same procedures as for Ft and Fc.

Dockerin-Cohesin Complex from C. cellulolyticum—The solution structure of the dockerin-cohesin complex from *C. cellulolyticum* was modeled using a three-step approach. First, the dockerin domain from *C. cellulolyticum* was generated with Swiss modeler (31) using the dockerin domain from *C. thermocellum* as template. The high identity (37%) between the dockerin sequences from *C. thermocellum* and *C. cellulolyticum* (Fig. 1) allows the assumption that they have a highly similar structure. Then the crystal structure of the cohesin from *C. cellulolyticum* (16) and the dockerin atomic model were positioned into the low resolution shape. Finally, the overall arrangement of the domains has been refined by molecular dynamics using the program CHARMM (32). While the cohesin of *C. cellulolyticum* was kept fixed, the dockerin domain was allowed to find its optimal position.

Normal Mode Analysis

We submitted the solution structures of Fc, Ft, Fccoh, and Ftcoh to the World Wide Web interface of the Elastic Network modeling (Elnémo) that provides a fast and simple tool to analyze low frequency normal modes of large protein assemblies (33). To analyze the overall motion of our proteins, we examined the B -factors derived from the mean square displacements, $\langle R^2 \rangle$, of all atoms in the 100 lowest frequency normal modes. The B -factors are computed using the relation $B = (8\pi^2/3) \langle R^2 \rangle$.

To detect differences between the domains, the B -factors calculated for each protein were normalized to unity at their minimal values.

RESULTS

Solution Structure of the Isolated Catalytic Module of Cel48F—Initially, we measured the solution scattering profile of the isolated catalytic module of Cel48F in order to check whether there might be any difference between the conformation in solution and the crystal structure, as proposed by

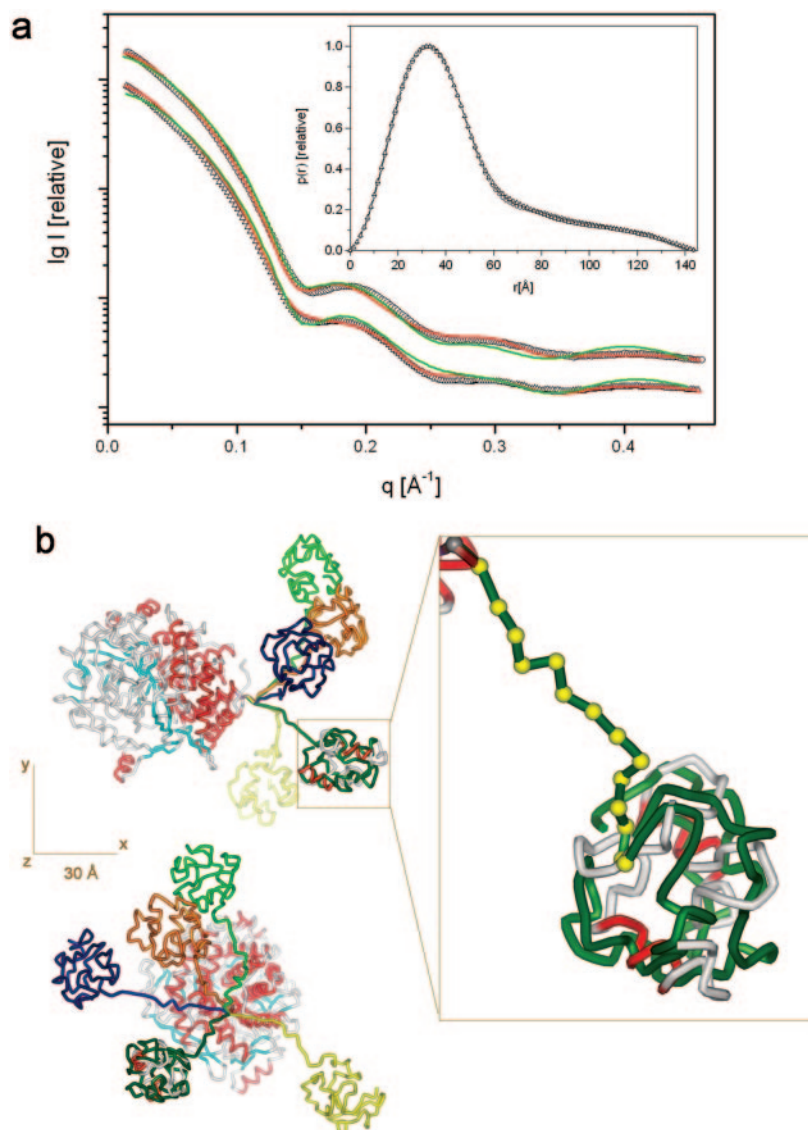
Parsiegla *et al.* (19). The experimental scattering data led to a radius of gyration of $R_G = 24.8 \pm 0.3$ Å. The distance distribution function $P(r)$ has a bell-shaped appearance that is typical of spherical molecules, with a maximum diameter D_{max} of 74 ± 2 Å (data not shown). We compared the experimental scattering profile with the calculated SAXS profile of the crystal structure (Protein Data Bank access code 1G9G). The calculated scattering profile is identical to the experimental one, fitting the data with a discrepancy χ of 1.1 and giving a theoretical radius of gyration $R_G = 24.7$ Å similar to the experimental value (Table I). These results suggest that the overall structure of isolated catalytic module of Cel48F in solution is identical to that in the crystal structure.

Solution Structure of the Free Entire Cellulase Cel48F—We studied two engineered cellulases: full-length Cel48F wild type from *C. cellulolyticum*, called henceforward Fc, corresponding to the catalytic module tethered through a linker to its dockerin; and the catalytic module of Cel48F covalently attached to the linker and the dockerin from *C. thermocellum* (see “Experimental Procedures”) that we called Ft (Fig. 1). The respective values of the radius of gyration $R_G = 34.9 \pm 0.8$ Å (Fc) and $R_G = 36.0 \pm 0.9$ Å (Ft), as deduced from the Guinier plots, are substantially increased with respect to the theoretical value ($R_G = 28$ Å) for a globular protein containing the same number of amino acids (34). This suggests that the proteins possess an extended conformation. The profiles of the $P(r)$ functions are highly similar between the two proteins (Fig. 2a, inset) and exhibit a long tail with a maximal particle dimension of $D_{\text{max}} = 142 \pm 6$ Å for Fc and $D_{\text{max}} = 147 \pm 4$ Å for Ft (Table I). The R_G values derived from the final distance distribution function $P(r)$ are in good agreement with the R_G calculated from the Guinier approximation. The shape of the $P(r)$ functions and the values observed for Fc and Ft provide evidence for high structural anisotropy of the particles (35) and also suggest that the two proteins adopt an extended conformation.

The overall shapes of Fc and Ft were calculated, and repetitive runs for each construct yielded superimposable models with similar overall structure and similar goodness of fit ($\chi \sim 1.0$; Table I). The overall shapes displayed a smaller and a larger globular unit connected by a stretched narrow region. The atomic structures of the dockerin determined by NMR (36) and of the catalytic module match with the size and shape, respectively, of each unit (data not shown).

Knowing the atomic structure of the catalytic module of Cel48F prompted us to apply the program package CREDO (28) to model the unknown linker/dockerin region from the experimental scattering curve obtained for the entire proteins. The models constructed in different, independent runs led to a fit with a discrepancy χ of ~ 2.2 and 2.0 for Fc and Ft, respectively. All models exhibit a stretched region of the same length, following the catalytic module in sequence, and ended in a small folded globular region. These models differ from each

FIG. 2. Solution structure of free entire cellulase Cel48F. *a*, experimental SAXS profiles of Fc (○) and Ft (△) fitted by the averaged form factors of the CREDO models obtained by the program OLIGOMER (red line) and SAXS profiles calculated from the final atomic models (green line). *Inset*, distance distribution functions of Fc (○) and Ft (△) computed from experimental SAXS profiles. The $P(r)$ functions are normalized to unity at their maximum. *b*, five typical CREDO models of linker-dockerin domains of Ft in a C_α tube representation displayed in different colors together with the crystal structure of catalytic module of Cel48F. The secondary structural elements of the crystal structures are in a C_α tube representation with cyan β -sheets and red α -helices. The bottom panel shows the models of the upper panel rotated counterclockwise by 90° around the y axis. *Inset*, crystal structure of the dockerin domain represented by white C_α tubes with red α -helices overlaid with one of the CREDO models (dark green). The yellow spheres indicate the location of C_α atoms as modeled by CREDO in the linker region.



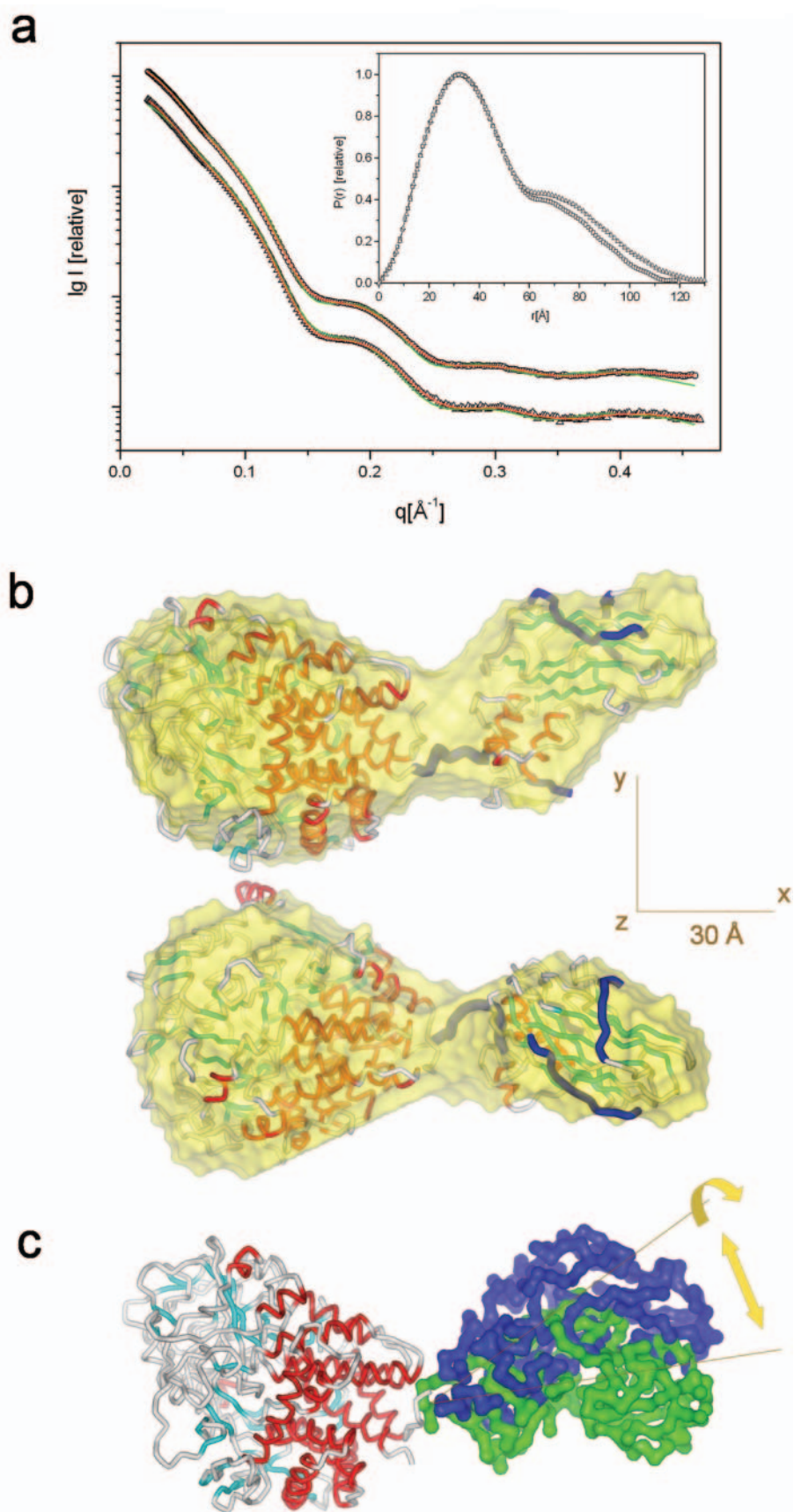
significantly with respect to the values of the free entire cellulases (Table I), although a complete module, comprising 170 residues, has been added. This shows that the complexed enzymes are markedly more compact than the free enzymes. This is further supported by the shape of the $P(r)$ functions, which are more typical of globular particles (35), and more strikingly by the D_{\max} values, which are significantly smaller ($D_{\max} = 122 \pm 3 \text{ \AA}$ for Fccoh, $D_{\max} = 128 \pm 4 \text{ \AA}$ for Ftcoh) (Table I) than for the free entire enzymes.

Fitting to the experimental data revealed the overall shapes with a discrepancy $\chi = 1.8$ for Fccoh and $\chi = 1.7$ for Ftcoh (Table I). The average shape was obtained using five independent shapes; averaging of independent reconstructions allows one to enhance the most persistent features of the bead models (27). An example of an average shape of Fccoh, presented in Fig. 3b, displays an ellipsoidal form with a smaller and a larger globular unit close to each other. The atomic structures of the individual domains were positioned into the average shape. The missing parts of the atomic structures, in particular the linker region, were modeled using *ab initio* modeling, which leads to extremely good discrepancy values to the experimental data ($\chi = 1.6$ for Fccoh, and $\chi = 1.7$ for Ftcoh). The restored fragments displayed slightly different conformations in independent modeling runs. However, they remain confined in the space bordered by the average shape (Fig. 3b). The models

including the missing fragments were used to construct the solution structure according to the procedure described under "Experimental Procedures." The scattering profiles fitting these solution structures to the experimental data are shown in Fig. 3a ($\chi = 1.6$ for Fccoh, and $\chi = 1.7$ for Ftcoh). The results of these model calculations show that the more compact overall shape of the complex arises from the pleated conformation of the linker between the catalytic module and the dockerin-cohesin region.

To exclude the possibility of any structural rearrangement in the dockerin-cohesin region, we compared the solution structure of the dockerin-cohesin region in the complex with the crystal structure of the isolated dockerin-cohesin. For this purpose, we used the program CREDO to calculate low resolution models of the dockerin-cohesin complex as observed in the SAXS experiments. Only the known crystal structure of the Cel48F catalytic module was used as a starting point to calculate the low resolution models of the linker, the dockerin, and the cohesin region. Independent runs provided similar models, with a χ of ~ 0.8 for Fccoh and χ of ~ 1.5 for Ftcoh, which only differ by their relative orientation. The superposition of several models suggests to some extent a conformational freedom in the linker. Fig. 3b depicts two border conformations limiting the possible movements. All shapes obtained are consistent with the form of the crystal structure of the isolated dockerin-

FIG. 3. Solution structure of the cellulases complexed with their cohesin. *a*, experimental SAXS profiles of Fccoh (○) and Ftcoh (△) fitted by the average form factors of the CREDO models, obtained by the program OLIGOMER (red line) and SAXS curves calculated from the final atomic models (green line). *Inset*, distance distribution functions of Fccoh (○) and Ftcoh (△) computed from experimental SAXS profiles. The $P(r)$ functions are normalized to unity at their maximum. *b*, superposition of the average overall shape of Fccoh calculated with GASBOR (yellow transparent surface) to the atomic model. The secondary structural elements of the crystal structure of catalytic module of Cel48F and the atomic model of the cohesin-dockerin module from *C. cellulolyticum* are in C_α tube representation with cyan β -sheets and red α -helices. The typical conformation of the constructed linker and missing loops using the program package CREDO are in a dark blue C_α tube representation. The *middle panel* shows the models of the *upper panel* rotated counterclockwise by 90° around the x axis. *c*, two restored models (green and blue) of the dockerin-cohesin complex of the Fccoh using the program CREDO. The CREDO models are displayed in surface representation. The *arrows* indicate the rearrangement of the dockerin-cohesin models between different runs.



cohesin complex (see below), and the different conformations are due to differences in the linker region. Independent single models displayed deviations from a best fit, and improvement of the fit was only achieved by OLIGOMER ($\chi = 0.6$ for Fccoh

and $\chi = 1.3$ for Ftcoh) (Fig. 3a, Table I), using the same approach as described above, assuming a multicomponent mixture of protein conformations. These results therefore show that the dockerin-cohesin domain probably undergoes rigid

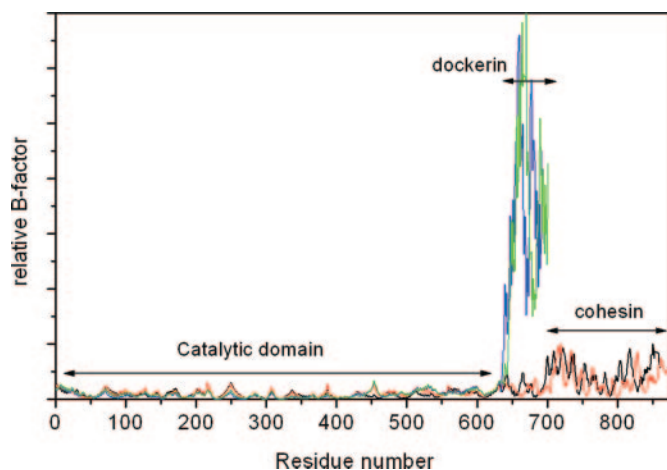


FIG. 4. **Normal mode analysis.** Residue mapping of atomic displacements in Fc (blue), Ft (green), Fccoh1c (black), and Ftcoh2t (red) presented as the B -factors calculated by Elnémo. The labels indicate the regions of the structure corresponding to the residue number. The B -factors are given in relative units.

body motion relative to the catalytic module but to a much less extent than the dockerin alone.

Normal Mode Analysis—To further explore the dynamic properties of the linker region in the presence of a free dockerin or bound to the cohesin, we supplemented our study with normal mode analysis (NMA). NMA is a powerful tool that provides information on the preferential direction of collective movement occurring during large rearrangements in molecular assemblies (37). Large scale rearrangements of the protein are often well represented by a small number of the lowest frequency normal modes (38). The protein movement can be represented as a superposition of the normal modes, fluctuating around a minimum energy conformation. To confirm the assumption that the free dockerin as well as the dockerin-cohesin complex undergoes rigid body motion relative to the catalytic module, the final models of the solution structure were subjected to NMA for both Fc and Ft in the absence or presence of their cognate cohesin. Theoretical B -factors for each atom were computed from the mean square displacements according to the 100 lowest frequency normal modes.

Fig. 4 shows that the relative B -factors of the catalytic module are identical in all constructs, with significantly lower values compared with those in the zones corresponding to the dockerin-cohesin region. Moreover, the linker and dockerin module exhibit much higher B -factor values in free Fc and Ft than in the complexed state with their respective cognate cohesins. NMA clearly indicates that the dockerin regions in Fc and Ft are able to undergo large displacements, whereas binding to the cognate cohesin restrains this movement. These results further confirm the results arising from the SAXS analysis.

Solution Structure of the Dockerin-Cohesin Complex—A more detailed analysis was performed on the docking region between the dockerin and the cohesin based on the SAXS data and the crystal structures. Five of the previous independent CREDO models of the dockerin-cohesin region were averaged. Fig. 5 shows that these low resolution models of the dockerin-cohesin complexes are highly similar for the *C. thermocellum* and the *C. cellulolyticum* variants. However, the low resolution model of the *C. thermocellum* variant (Fig. 5, left panel) has a somewhat bulkier appearance, indicative of a larger conformational freedom (39). This might be due to its linker between the dockerin and the catalytic module that is longer by 7 residues.

We superimposed the low resolution model of the dockerin-cohesin complex of *C. thermocellum* obtained by SAXS with its

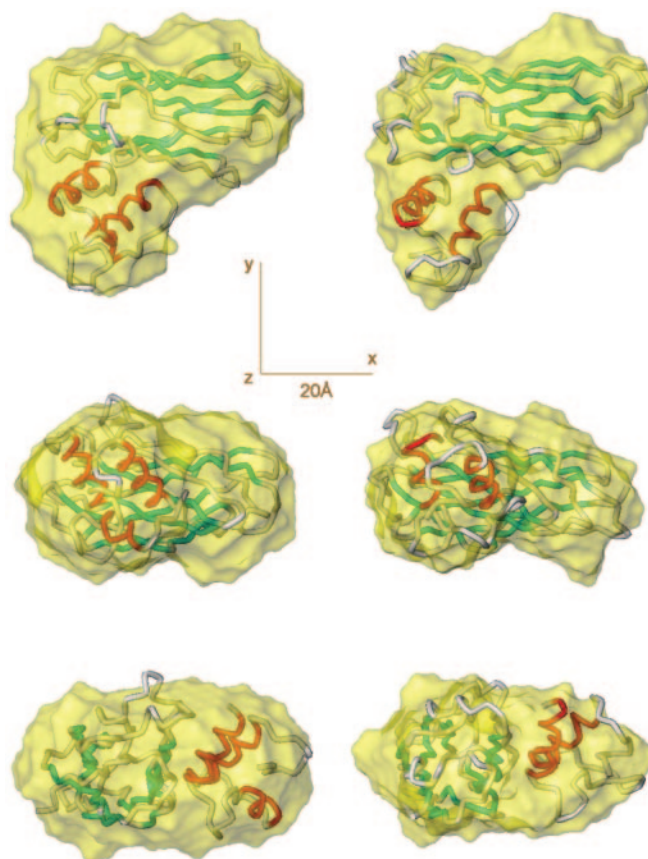


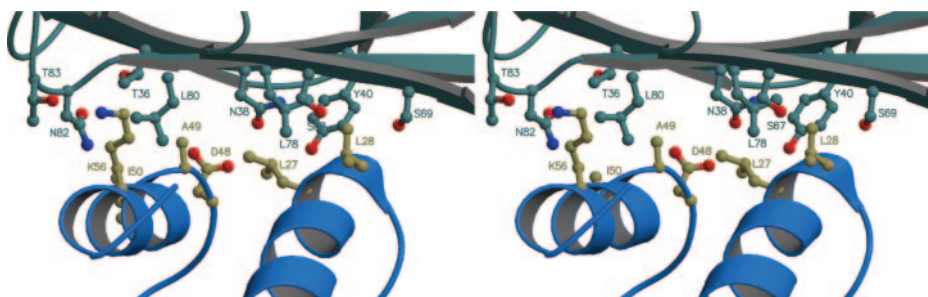
FIG. 5. **Solution structures of dockerin-cohesin complexes from *C. thermocellum* and *C. cellulolyticum*.** Average low resolution models of cohesin-dockerin complexes from *C. thermocellum* (left column) and *C. cellulolyticum* (right column) displayed in a yellow transparent surface representation superimposed with the crystal structure of the cohesin-dockerin complex from *C. thermocellum* (left column) and the atomic model from *C. cellulolyticum* (right column). The atomic structures are displayed in a C_{α} tube representation with cyan β -sheets and red α -helices. In both panels, the middle and bottom rows are rotated clockwise by 90° around the x and z axis, respectively.

known crystal structure. The low resolution model superimposes well with the atomic structure and shows that the overall organization of the dockerin-cohesin complex of *C. thermocellum* in solution is identical to that in the crystal. The similarity of both complexes (a direct consequence of the high sequence similarity) incited us to produce an atomic model of the dockerin-cohesin complex of *C. cellulolyticum*, based on the existing crystal structure of the cohesin from the latter and a sequence-based model of the dockerin, guided by the crystal structure of the dockerin-cohesin complex of *C. thermocellum* and the low resolution shape, obtained from the SAXS experiments. This atomic model, constructed as described under “Experimental Procedures,” is shown superimposed on the low resolution model of the dockerin-cohesin complex of *C. cellulolyticum* (Fig. 5, right panel). One should note that the well documented internal 2-fold symmetry of the dockerin (17) would lead to two equivalent orientations that cannot be distinguished with this procedure. The model thus obtained clearly exhibits a similar arrangement to that of the crystal structure of the complex of *C. thermocellum*.

DISCUSSION

We have established the structural determinants of a cellulosomal cellulase appended with a dockerin in solution and of the complex formed with the cognate cohesin by joining small angle scattering studies with the known atomic structures of

FIG. 6. Atomic model of the interface between the dockerin and the cohesin from *C. cellulolyticum*. The residues involved in domain contacts are shown as ball-and-stick models colored in blue-green and dark khaki for the cohesin and the dockerin, respectively.



the isolated modules. The work reported here demonstrates the advantages of combining these complementary techniques. SAXS is indeed a very useful tool to investigate the global structure of protein complexes in solution (40). The intrinsic limitation of this technique to a low resolution can be overcome by the combined use of data at the atomic scale provided for example by x-ray diffraction. This offers a unique alternative method to electron microscopy that enables the study of lower molecular weight assemblies and that helps to circumvent the impediments of high resolution techniques, such as the failure to crystallize certain proteins due to flexibility and due to heterogeneity in the primary, tertiary, and/or quaternary structure, etc. (39). The application of this procedure in our study permitted us not only to investigate the overall shapes of the complexes in solution but also to propose a model of the *C. cellulolyticum* dockerin-cohesin at the atomic scale. It is, however, important to remember that the scattering from dissolved particles yields an average over the ensemble in the irradiation volume, and if the particles are flexible, modeling of SAXS data provides an average conformation in solution (20, 41, 42).

We performed SAXS experiments on full-length cellulase Cel48F containing the native *C. cellulolyticum* dockerin and on an engineered cellulase appended with a *C. thermocellum* dockerin as well as their assemblies with the cognate cohesin. The models obtained for the full-length cellulases indicate that the dockerin domain is folded and adopts a conformation identical to that of the isolated dockerin determined by NMR (36) (Fig. 2b). The models inferred from the SAXS data and results from NMA analysis also show that the linker connecting the catalytic module to the dockerin is extended and flexible (Fig. 4). On the other hand, the results obtained for the protein assemblies clearly define a more compact character for the complex, with a lower flexibility. This implies that docking of the full-length cellulase to the cohesin leads to the pleating of the cellulase linker, although some flexibility persists.

These data are consistent with a general feature of the cellulosomal enzymes from *Clostridia* produced in *Escherichia coli*, which undergo spontaneous cleavage in the linker region upon purification or storage above 0 °C (43, 44). It has not been established yet whether the cleavage is due to nonproteolytic cleavage of aspartyl bonds in the unfolded stretch (45, 46) or to the co-purification of trace amounts of proteases. Nevertheless, it is generally observed that upon complexation of the recombinant enzymes onto cohesins or scaffoldins, cleavage no longer occurs (12). In this context, the observed compaction of the enzyme linker upon binding to the cohesin suggests that the structuring of the linker prevents the cleavage of an aspartyl bond or is no longer accessible to residual proteases.

The quality of our data together with the atomic structure of the *C. thermocellum* counterpart of the cohesin-dockerin complex allowed the construction of an atomic model of the docking interaction in *C. cellulolyticum*. A close examination of the interface of the dockerin-cohesin model (Fig. 6) shows that the residues of the cohesin that are interacting match those predicted by Spinelli *et al.* (16) (Thr³⁶, Asn³⁸, Tyr⁴⁰, Ser⁶⁷, Ser⁶⁹,

Leu⁷⁸, Leu⁸⁰, Asn⁸², and Thr⁸³). On the side of the dockerin, the following residues are in direct contact with the cohesin: Leu²⁷, Leu²⁸, Asp⁴⁸, Ala⁴⁹, and Lys⁵⁶. The sequence alignment in Fig. 1 highlights the sequence location of these residues as compared with those of the complex of *C. thermocellum*. Whereas the nature and overall location of the interacting regions in the two dockerin-cohesin complexes are similar, the details of interaction show significant differences. Indeed, Leu²⁸ is in direct contact distance to Tyr⁴⁰ in the *C. cellulolyticum* complex, and the hydrogen bond across the interface is comparable with that observed in the complex of *C. thermocellum* (17) (Leu²²-Tyr⁷⁴); however, the relative position of the respective tyrosines are quite different. In the *C. cellulolyticum* model, the dockerin residue Lys⁵⁶ forms a hydrogen bond with Thr⁸³. In the same location, a salt bridge interaction has been observed in the complex of *C. thermocellum* involving Arg⁵³ (dockerin) and Glu⁸⁶ of the cohesin. Furthermore, residues Ala⁴⁹ and Ile⁵⁰ of the *C. cellulolyticum* dockerin model (which are serine and threonine in the dockerin of *C. thermocellum*) are involved in the interface, which is consistent with the study of Mechaly *et al.* (47), who showed by mutagenesis that these residues play a key role in the specific dockerin-cohesin interactions. As proposed by Jindou *et al.* (48), Ala⁴⁹-Ile⁵⁰ are not the only key residues responsible for the species specificity but rather allow the positioning of the other residues at the right location with respect to the corresponding residues of the cohesin. Finally, Asn³⁸ of the cohesin in *C. cellulolyticum*, replaced by an aspartate in *C. thermocellum*, is located at the cohesin/dockerin interface in our model and represents another key position for species specificity. This is consistent with a site-directed mutagenesis study in *C. thermocellum* (Asp to Asn), which drastically reduced the affinity for the dockerin (49). It is worthwhile noting that residues Ala⁴⁹ and Ile⁵⁰ of the *C. cellulolyticum* dockerin are highly conserved in the different *C. cellulolyticum* dockerins. Similarly, Asn³⁸ of the cohesin belongs to a stretch, Thr-Cys-Asn-Phe-Tyr, that is conserved in all of the cohesins from *C. cellulolyticum*, even in the most divergent ones. Even more interestingly, the same stretch is found on all six cohesins from *Clostridium josui* (50), and incidentally, the dockerins of the latter also bear the sequence Ala-Ile/Leu. On this basis, cohesins and dockerins from these two species are suspected to cross-react; this has, however, never yet been experimentally demonstrated. Summarizing these observations, we can conclude that our model, based on the global form calculated according to the SAXS data, is in good agreement with all experimental data reported to date, and that, although it remains a model, it helps to identify the factors governing species specificity of the dockerin-cohesin complexes in *C. thermocellum* versus *C. cellulolyticum*. It appears from our model that simple point mutations of residues in the dockerin sequence can establish a new specificity for a cohesin from a different species, having a docking site with the complementary nature. In other words, the overall organization of the dockerin-cohesin complex is the same for both organisms, and simply local differences in the sequences are most

likely responsible for the specificity of the docking. We intend to verify this assumption in future experimental work.

Apparently, the species-specific recognition of the cognate partners initiates a structuring of the entire enzyme when it is incorporated into cellulosomes. Hence, the anchoring of cellulases into natural or artificial cellulosomes triggers enhanced synergy between the catalytic subunits (5, 12). In this respect, one can hypothesize that the pleating of the enzyme linker upon binding may induce a closer proximity between the catalytic modules within the cellulosomes, thus leading to optimal cooperativity for the degradation of crystalline cellulose and prevention of any steric clashes between adjacent catalytic modules that would ruin this cooperativity. The residual flexibility also allows short scale motion to adjust the respective enzyme positions on the substrate and provides an explanation of the enhanced synergy between the enzymes, since a rigid conformation of each catalytic subunit in the cellulosomes would prevent improved cooperativity. Moreover, the resolution of the calculated shapes (13 Å) enabled localization of the active site cleft opposite to the dockerin-cohesin assembly. Consequently, when the cellulosome assembles, the catalytic sites come to lie on the external surface of this multiprotein complex.

Similar studies performed on the bimodular cellulase Cel45 from *Humicola insolens* (42) also explored the role of the linker in the synergy of the enzyme. This fungal enzyme, which does not form complexes *in vivo*, contains a catalytic module and a cellulose binding module (CBM) separated by a glycosylated linker peptide of 36 residues. SAXS analysis indicated that the linker has an extended and flexible conformation, probably allowed by its transversal *O*-glycosylation. A model where the cellulase moves on cellulose following a caterpillar-like motion was thus proposed to explain the enhanced activity of the bimodular enzyme with respect to the isolated catalytic module; whereas the CBM remains bound to a specific site, the catalytic module hydrolyzes several glycosidic bonds until the linker eventually becomes too compressed, inducing the translation of the CBM along the cellulose surface. In the case of cellulosomal enzymes, this model cannot be applied, since they usually do not contain any genuine CBM, and the anchoring to the substrate is mediated by the powerful family 3A CBM of the scaffoldin. The linker peptides of Fc and Ft, which also display an extended conformation in the free state of the enzymes, are not glycosylated, and they adopt a compact conformation upon binding of the cellulases to the cognate cohesins. Although some flexibility of the linker is still permitted in the complexed state, a dramatic extension of the linker peptide in the presence of cellulose seems unlikely to occur. Besides, the linker peptides joining catalytic modules to dockerins in *Clostridia* are rather small (10–12 residues) compared with bimodular cellulase linkers separating the catalytic module and the CBM (often several dozen residues and up to hundreds and more). All of these observations converge toward the suggestion that a high flexibility is not required between the catalytic module and the cognate cohesin for the synergy of the complex. On the other hand, in clostridial scaffoldins, which were found to be glycosylated (51, 52), the CBM and the cohesins are separated by Ser/Thr-containing linker peptides, suspected to be *O*-glycosylated (51). Similarly, linkers joining two different modules on the scaffolding can be rather long (up to 35 residues). Thus, whereas the model proposed for Cel45 from *Humicola insolens* cannot be applied to individual cellulosomal enzymes, one may hypothesize that such processes may be relevant for the entire cellulosome particle. Cellulosome synergy would be thus allowed through an accordion-like motion of the scaffoldin that would result from the flexibility of each intermodule linker, enabling an adjustment of the catalytic

activity according to the local and global geometric requirements of the substrate. Further, small angle scattering studies on cellulosomal constructions mimicking the multiprotein complex, relying on the work reported here, would help to address this question and are currently under way.

Acknowledgments—We thank Odile Valette for expert technical assistance and Maxim Petoukhov for fruitful discussions of the structural modeling aspects.

REFERENCES

- Bayer, E. A., Bélaïch, J.-P., Shoham, Y., and Lamed, R. (2004) *Annu. Rev. Microbiol.* **58**, 521–524
- Adams, J. (2003) *Cancer Treat. Rev.* **29**, Suppl. 1, 3–9
- Wilson, D. N., and Nierhaus, K. H. (2003) *Angew. Chem. Int. Ed. Engl.* **42**, 3464–3486
- Fierobe, H. P., Pages, S., Belaich, A., Champ, S., Lexa, D., and Belaich, J. P. (1999) *Biochemistry* **38**, 12822–12832
- Fierobe, H. P., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J. P., and Bayer, E. A. (2001) *J. Biol. Chem.* **276**, 21257–21261
- Yaron, S., Morag, E., Bayer, E. A., Lamed, R., and Shoham, Y. (1995) *FEBS Lett.* **360**, 121–124
- Pages, S., Belaich, A., Fierobe, H. P., Tardif, C., Gaudin, C., and Belaich, J. P. (1999) *J. Bacteriol.* **181**, 1801–1810
- Perret, S., Belaich, A., Fierobe, H. P., Belaich, J. P., and Tardif, C. (2004) *J. Bacteriol.* **186**, 6544–6552
- Maamar, H., Valette, O., Fierobe, H. P., Belaich, A., Belaich, J. P., and Tardif, C. (2004) *Mol. Microbiol.* **51**, 589–598
- Gerngross, U. T., Romaniec, M. P., Kobayashi, T., Huskisson, N. S., and Demain, A. L. (1993) *Mol. Microbiol.* **8**, 325–334
- Pages, S., Belaich, A., Belaich, J. P., Morag, E., Lamed, R., Shoham, Y., and Bayer, E. A. (1997) *Proteins* **29**, 517–527
- Fierobe, H. P., Bayer, E. A., Tardif, C., Czjzek, M., Mechaly, A., Belaich, A., Lamed, R., Shoham, Y., and Belaich, J. P. (2002) *J. Biol. Chem.* **277**, 49621–49630
- Shimon, L. J., Pages, S., Belaich, A., Belaich, J. P., Bayer, E. A., Lamed, R., Shoham, Y., and Frolow, F. (2000) *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1560–1568
- Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) *EMBO J.* **15**, 5739–5751
- Shimon, L. J., Bayer, E. A., Morag, E., Lamed, R., Yaron, S., Shoham, Y., and Frolow, F. (1997) *Structure* **5**, 381–390
- Spinelli, S., Fierobe, H. P., Belaich, A., Belaich, J. P., Henrissat, B., and Cambillau, C. (2000) *J. Mol. Biol.* **304**, 189–200
- Carvalho, A. L., Dias, F. M., Prates, J. A., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M., Romão, M. J., and Fontes, C. M. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13809–13814
- Svergun, D. I., and Koch, M. H. (2002) *Curr. Opin. Struct. Biol.* **12**, 654–660
- Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (1998) *EMBO J.* **17**, 5551–5562
- Hammel, M., Krichbaum, M., Gries, A., Kostner, G. M., Laggner, P., and Prassl, R. (2002) *J. Mol. Biol.* **321**, 85–97
- Svergun, D. (1992) *J. Appl. Crystallogr.* **25**, 495–503
- Bergmann, A., Fritz, G., and Glatter, O. (2000) *J. Appl. Crystallogr.* **33**, 1212–1216
- Svergun, D., Barabero, C., and Koch, M. H. (1995) *J. Appl. Crystallogr.* **28**, 768–773
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) *J. Appl. Crystallogr.* **36**, 1277–1282
- Svergun, D. I. (1999) *Biophys. J.* **76**, 2879–2886
- Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) *Biophys. J.* **80**, 2946–2953
- Volkov, V. V., and Svergun, D. I. (2003) *J. Appl. Crystallogr.* **36**, 860–864
- Petoukhov, M. V., Eady, N. A., Brown, K. A., and Svergun, D. I. (2002) *Biophys. J.* **83**, 3113–3125
- Kozin, M. B., and Svergun, D. I. (2001) *J. Appl. Crystallogr.* **34**, 33–41
- Konarev, P. V., Petoukhov, M. V., and Svergun, D. I. (2001) *J. Appl. Crystallogr.* **34**, 527–532
- Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**, 2714–2723
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comp. Chem.* **4**, 187–217
- Suhre, K., and Sanejouand, Y. H. (2004) *Nucleic Acids Res.* **32**, W610–W614
- Millett, I. S., Doniach, S., and Plaxco, K. W. (2002) *Adv. Protein. Chem.* **62**, 241–262
- Glatter, O. (1982) in *Small Angle X-ray Scattering* (Kratky, O., ed) pp. 167–196, Academic Press, London
- Lytle, B. L., Volkman, B. F., Westler, W. M., Heckman, M. P., and Wu, J. H. (2001) *J. Mol. Biol.* **307**, 745–753
- Delarue, M., and Dumas, P. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6957–6962
- Tama, F., and Sanejouand, Y. H. (2001) *Protein Eng.* **14**, 1–6
- Marquez, J. A., Smith, C. I., Petoukhov, M. V., Lo Surdo, P., Mattsson, P. T., Knekt, M., Westlund, A., Scheffzek, K., Saraste, M., and Svergun, D. I. (2003) *EMBO J.* **22**, 4616–4624
- Koch, M. H., Vachette, P., and Svergun, D. I. (2003) *Q. Rev. Biophys.* **36**, 147–227
- Aslam, M., and Perkins, S. J. (2001) *J. Mol. Biol.* **309**, 1117–1138
- Receveur, V., Czjzek, M., Schulein, M., Panine, P., and Henrissat, B. (2002) *J. Biol. Chem.* **277**, 40887–40892

43. Fierobe, H. P., Bagnara-Tardif, C., Gaudin, C., Guerlesquin, F., Sauve, P., Belaich, A., and Belaich, J. P. (1993) *Eur. J. Biochem.* **217**, 557–565
44. Reverbel-Leroy, C., Pages, S., Belaich, A., Belaich, J. P., and Tardif, C. (1997) *J. Bacteriol.* **179**, 46–52
45. Chen, H. M., Ford, C., and Reilly, P. J. (1995) *Protein Eng.* **8**, 575–582
46. Lamed, R., Kenig, R., Morag, E., Yaron, S., Shoham, Y., and Bayer, E. A. (2001) *Appl. Biochem. Biotechnol.* **90**, 67–73
47. Mechaly, A., Fierobe, H. P., Belaich, A., Belaich, J. P., Lamed, R., Shoham, Y., and Bayer, E. A. (2001) *J. Biol. Chem.* **276**, 9883–9888
48. Jindou, S., Soda, A., Karita, S., Kajino, T., Beguin, P., Wu, J. H., Inagaki, M., Kimura, T., Sakka, K., and Ohmiya, K. (2004) *J. Biol. Chem.* **279**, 9867–9874
49. Handelsman, T., Barak, Y., Nakar, D., Mechaly, A., Lamed, R., Shoham, Y., and Bayer, E. A. (2004) *FEBS Lett.* **572**, 195–200
50. Kakiuchi, M., Isui, A., Suzuki, K., Fujino, T., Fujino, E., Kimura, T., Karita, S., Sakka, K., and Ohmiya, K. (1998) *J. Bacteriol.* **180**, 4303–4308
51. Gerwig, G. J., Kamerling, J. P., Vliegthart, J. F., Morag, E., Lamed, R., and Bayer, E. A. (1993) *J. Biol. Chem.* **268**, 26956–26960
52. Gal, L., Pages, S., Gaudin, C., Belaich, A., Reverbel-Leroy, C., Tardif, C., and Belaich, J. P. (1997) *Appl. Environ. Microbiol.* **63**, 903–909