Protein Science (1997), 6:294–303. Cambridge University Press. Printed in the USA.

Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from *Trichoderma reesei*

MAIJA-LIISA MATTINEN,[1] MAARIT KONTTELI,[1] JANNE KEROVUO,[1] MARKUS LINDER,[2] ARTO ANNILA,[1] GUNNAR LINDEBERG,[3] TAPANI REINIKAINEN,[2] AND TORBJÖRN DRAKENBERG[1]

[1]VTT, Chemical Technology, Box 1401, FIN-02044 VTT, Finland
[2]VTT, Biotechnology and Food Research, Box 1500, FIN-02044 VTT, Finland
[3]Department of Medical and Physiological Chemistry, University of Uppsala, Box 575, S-75124 Uppsala, Sweden

(RECEIVED July 31, 1996; ACCEPTED November 20, 1996)

Abstract

Three-dimensional solution structures for three engineered, synthetic CBDs (Y5A, Y31A, and Y32A) of cellobiohydrolase I (CBHI) from *Trichoderma reesei* were studied with nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy. According to CD measurements the antiparallel β-sheet structure of the CBD fold was preserved in all engineered peptides. The three-dimensional NMR-based structures of Y31A and Y32A revealed only small local changes due to mutations in the flat face of CBD, which is expected to bind to crystalline cellulose. Therefore, the structural roles of Y31 and Y32 are minor, but their functional importance is obvious because these mutants do not bind strongly to cellulose. In the case of Y5A, the disruption of the structural framework at the N-terminus and the complete loss of binding affinity implies that Y5 has both structural and functional significance. The number of aromatic residues and their precise spatial arrangement in the flat face of the type I CBD fold appears to be critical for specific binding. A model for the CBD binding in which the three aligned aromatic rings stack onto every other glucose ring of the cellulose polymer is discussed.

Keywords: cellulase; nuclear magnetic resonance spectroscopy; protein–carbohydrate interaction; structure determination

In nature, cellulose is degraded by a variety of micro-organisms. The filamentous fungus *Trichoderma reesei* has been widely used as a model organism in studies of cellulose breakdown (Quiocho, 1986, 1993; Claeyssens & Tomme, 1989; Vyas, 1991; Poole et al., 1993; Din et al., 1994). *T. reesei* secretes a mixture of cellulases consisting of two cellobiohydrolases (CBHI and CBHII) and at least three endoglucanases (EGI, EGII, and EGV), which have been shown to act synergistically (Henrissat et al., 1985; Wood & Garcia-Campayo, 1990; Irwin et al., 1993; Nidetzky et al., 1993; Béguin & Aubert, 1994). These enzymes have a catalytic core domain and a type I cellulose-binding domain (CBD), which are connected by a glycosylated linker peptide.

Much of the current knowledge on cellulase–cellulose binding stems from separate studies of the soluble enzyme and the solid substrate because the complete biphasic system is difficult to investigate directly.

The core domains of both CBHI and CBHII are thought to degrade solubilized chains by attacking the free ends and subsequently releasing cellobiosyl units. EGs hydrolyse amorphous microcrystalline cellulose by cleaving the polymer chains at random. β-glucosidases complete the hydrolytic process yielding glucose to be metabolized by the organism (Cambillau & van Tilbeurgh, 1993).

The catalytic core has some binding affinity for crystalline cellulose in the absence of even CBD, but it is much lower than that of the intact enzyme, whereas detached CBD alone binds strongly to cellulose. Thus, CBD is required for the full activity of cellulases on crystalline cellulose (van Tilbeurgh et al., 1986; Ståhlberg et al., 1988, 1991; Tomme et al., 1988; Reinikainen et al., 1992,

Reprint requests to: Maija-Liisa Mattinen, VTT, Chemical Technology, Box 1400, FIN-02044 VTT, Finland; e-mail: maija.mattinen@vtt.fi.

Abbreviations: 2D, two-dimensional; CBD, cellulose-binding domain; CBD_{CBHI}, type I CBD of CBHI; CBD_{Cex}, type II CBD of Cex; CBD_{EGI}, type I CBD of EGI; CBH, cellobiohydrolase; CBHI, cellobiohydrolase I; CBHII, cellobiohydrolase II; CD, circular dichroism; Cex, exoglucanase/xylanase from *Cellulomonas fimi*; COSY, correlation spectroscopy; DG, distance geometry; EG, endoglucanase; EGI, endoglucanase I; EGII, endoglucanase II; EGV, endoglucanase V; FID, free induction decay; MD, molecular dynamics; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; RELAY-COSY, relayed coherence transfer correlation spectroscopy; SA, simulated annealing; TOCSY, total correlation spectroscopy.

1995), however, the detailed binding mechanism of the CBD is not known.

The three-dimensional structure of CBD_{CBHI} has been determined by NMR spectroscopy (Kraulis et al., 1989). The main secondary structure of CBD_{CBHI} is an irregular anti-parallel triple-stranded β-sheet. The peptide folds into a wedge-shaped structure. One face of the wedge is flat and hydrophilic, while the other face is rougher and less hydrophilic. Although both faces contain segments of conserved amino acids, it is the flat face with three aromatic residues that is presumed to bind to cellulose (Linder et al., 1995a, 1995b; Reinikainen et al., 1995). This conclusion has been supported by point mutation experiments. When the tyrosines on the flat face were replaced one by one by alanine, the binding affinity to cellulose was dramatically reduced (Linder et al., 1995a). The two amide groups neighboring the tyrosines were also shown to contribute to the binding strength (Linder et al., 1995a); however, the structural consequences induced by the mutations remained obscure.

In the present work, the structural changes caused by the point mutations (Y5A, Y31A, and Y32A) were examined by determining the three-dimensional structures of the peptides.

Results

CD spectroscopy

Qualitatively the CD spectra of the mutants resemble that of the wild type, showing that the secondary structures of the engineered peptides are similar to those of the wild type (Fig. 1). All spectra have a band in the region around 205–250 nm, which is characteristic of an anti-parallel β-sheet. Otherwise, the spectra are not typical for a β-sheet structure due to short strands and many type I turns in the CBD fold. The low amplitude and merely qualitative spectral characteristics do not allow any detailed conclusions to be drawn regarding structural changes in the engineered peptides as compared to the wild type.

NMR spectroscopy

As has been shown, earlier chemical shifts of the mutants differed relatively little from shifts of the wild type (Linder et al., 1995a).

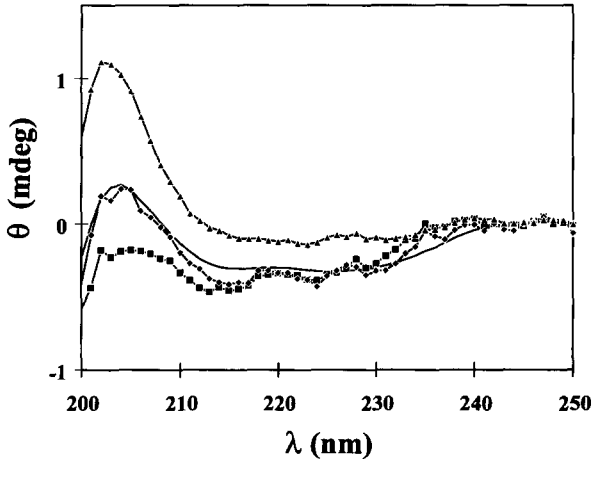

Fig. 1. Near-UV CD spectra of Y5A (diamond), Y31A (square), Y32A (triangle), and wild type (—). The spectrum of the background (solvent) has been subtracted from the spectra of the samples.

This is, however, not necessarily an indication of small structural changes because in a small domain like CBD a majority of ^1H shifts do not deviate strongly from the random coil shifts. In general, the NMR spectra of Y31A and Y32A are very similar to those of the wild type. For Y5A, certain resonances are clustered and the number of cross-peaks in the NOESY spectra is significantly smaller (Fig. 2).

Three-dimensional structures

From the final sets of calculated structures, only those with at most two violations of no more than 0.4 Å were accepted for subsequent examination (Fig. 3). For Y5A, 37 structures were accepted. The average of the calculated energies for these structures is about 70 kcal/mol. For Y31A and Y32A, using the same selection criteria as for Y5A, 18 and 19 structures were accepted, respectively. The average of the calculated energies for these structures is about 20 kcal/mol. Qualitatively for the less-ordered structure of Y5A the calculated energy is larger than for the well-structured Y31A and Y32A. The calculated energy for the wild type is even smaller. The number of violated distance constraints is comparable for all of the engineered peptides (Table 1).

A comparison of the engineered CBDs with the wild type shows an unmistakable resemblance. The resolution of Y31A and Y32A allows a close examination. For Y31A and Y32A the rms deviations of the backbone structures (0.53 Å for Y31A and 0.81 Å for Y32A) are only slightly larger than that of the wild type (0.33 Å) (Table 1). The side chains are determined to a resolution of about 1 Å. The structures of Y5A have on an average a clearly inferior resolution (~2 Å) for the backbone atoms and also larger dispersion for the side chains (~2.6 Å). For all the engineered peptides, the rms deviation per residue correlate well with the number of NOEs per residue (Fig. 4). For Y31A and Y32A, the rms deviation is small and fluctuates by about 1 Å along the whole sequence, whereas for Y5A it vary much more along the sequence. The deviation is largest in the N-terminus and between residues G15–C25. Even the relative position of the N-terminus with respect to loop S14–C19 remains uncertain. Consequently, the structure generation resulted in two sets of families one for each topological possibility. N-terminus is either above or below the plane of the turn. The segment from Q7 to G15 and Q26 to L36, on the other hand, have comparatively small (~2 Å) rms deviations. For all the engineered peptides the β-sheets are better defined than the loops and turns.

Flat face

The structural consequences of mutations to the flat face are seen in Figures 5 and 6 where the backbones of the engineered peptides are superimposed on the wild type. Only the backbone trace of the structure with the smallest rms deviation with respect to the wild type is shown for clarity. In Figure 5 the side chains of the residues on the flat face (5, 31, 32, 29, 34) are drawn for all members of the families to reveal the resolution. The resolution of the wild-type structure is comparable; however, the side chains are shown for only one of these structures.

The flat face of Y31A is quite well preserved. A31 has moved only slightly towards N29 as compared to Y31, and there are only minor rearrangements in the remaining residues of the flat face compared to the wild type. The aromatic rings of Y5 and Y32 occupy slightly different positions compared to the wild type. Y5

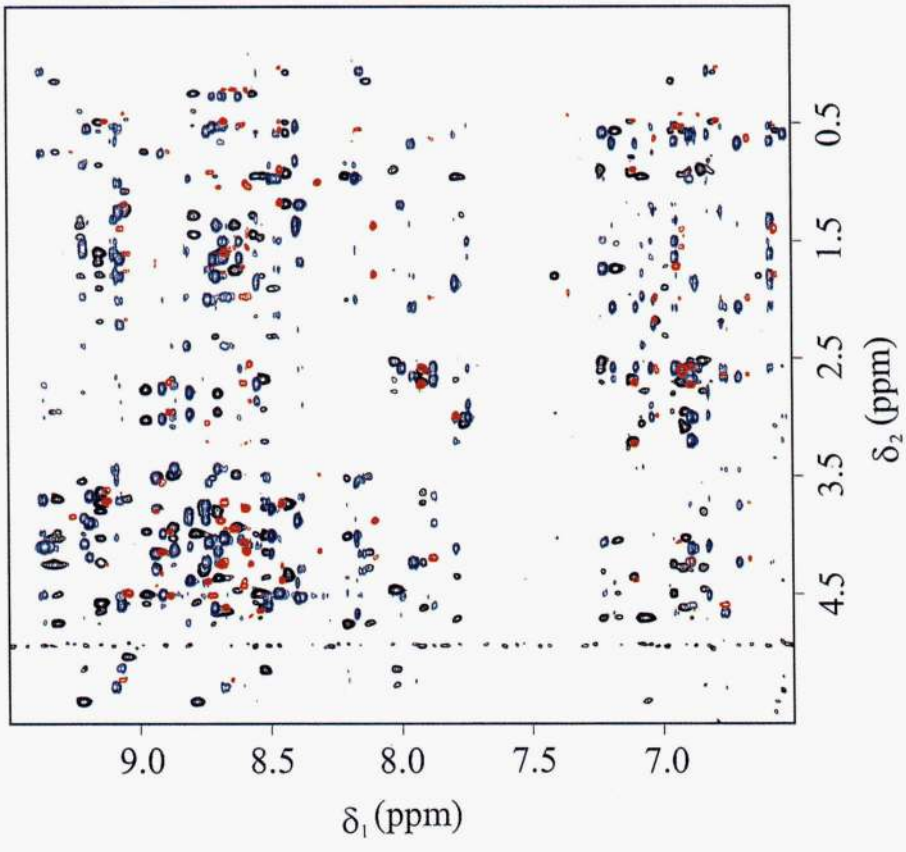

Fig. 2. NH (δ_1 axis)-aliphatic (δ_2 axis) regions of the 2D 1H-1H NOESY spectra of the engineered peptides (red: Y5A, blue: Y31A, and black: Y32A). The spectra are superimposed to facilitate comparison. Note that the wild type is very similar to Y31A and therefore not included.

is tilted about 30 degrees from the plane of the flat face. N29 and Q34 are in their wild-type–like positions within the precision of the structure.

In the case of Y32A, the flat face has become concave and A32 is buried in the interior. Due to the mutation, the residues that neighbor Y32 in the wild-type structure have obtained more space. The amide groups of Q34 and N29 have moved closer to each other filling out the space left by Y32. Also, Y31 and Y5 have moved towards each other by about 5 Å and their rings have tilted out of the flat face of the wild type. Thus, this engineered CBD has neither the planarity nor the periodicity of the residues in the flat face of the wild type.

In the case of Y5A the flat face is less precisely defined compared to the other engineered CBDs, because there is too little experimental data to determine the position of the N-terminus relative to the plane of the loop S14–C19. When the N-terminus is below the plane, A5 is approximately in the plane of the flat face, whereas in the other case it is far above. For both families of structures the remaining residues of the flat face have average positions similar to those in the wild type.

Discussion

Structural framework

The CBD_{CBHI} binds on slightly acidic conditions, and the pH dependence is weak (Reinikainen et al. 1995). The compact triple-stranded anti-parallel β-sheet of the type I CBD fold constitutes a framework for the functionally important residues on the flat face. The backbones of the engineered CBDs Y31A and Y32A follow the fold of the wild type (Fig. 3). The strands and turns are mainly preserved. Furthermore, the backbones are well defined along the entire sequence (Fig. 4), and most of the side-chain positions are well defined for these peptides. This implies that the mutations of Y31 and Y32 neither alter nor decrease the stability of the CBD fold, i.e., they are not part of the structural framework.

In contrast, the mutation of Y5 to alanine has obvious structural consequences. Although segments that correspond to the anti-parallel β-strands and the connecting turns remain rather well defined, the N-terminus as well as the spatially close segment G15–C25 have poor definitions. The structure definition is for this part of the peptide so poor that the mutual positions of the N-terminus and the β2-strand (residues 24–28) remain ambiguous. For Y5A, approximately half of the accepted structures have the N-terminus above the loop 14–19 as the wild type, whereas the other half have it below. In general, Y5A has less order than Y31A and Y32A. This may imply that the overall stability of the structure is compromised. It is concluded that Y5 plays a key role in maintaining the structural integrity.

To ensure reliability of the Y5A results, 17 additional DGII structures were calculated without any ϕ-angle restraint for A5. It was suspected that this restraint determines the location of the N-terminus due to the probable structural role of tyrosine 5. Nev-

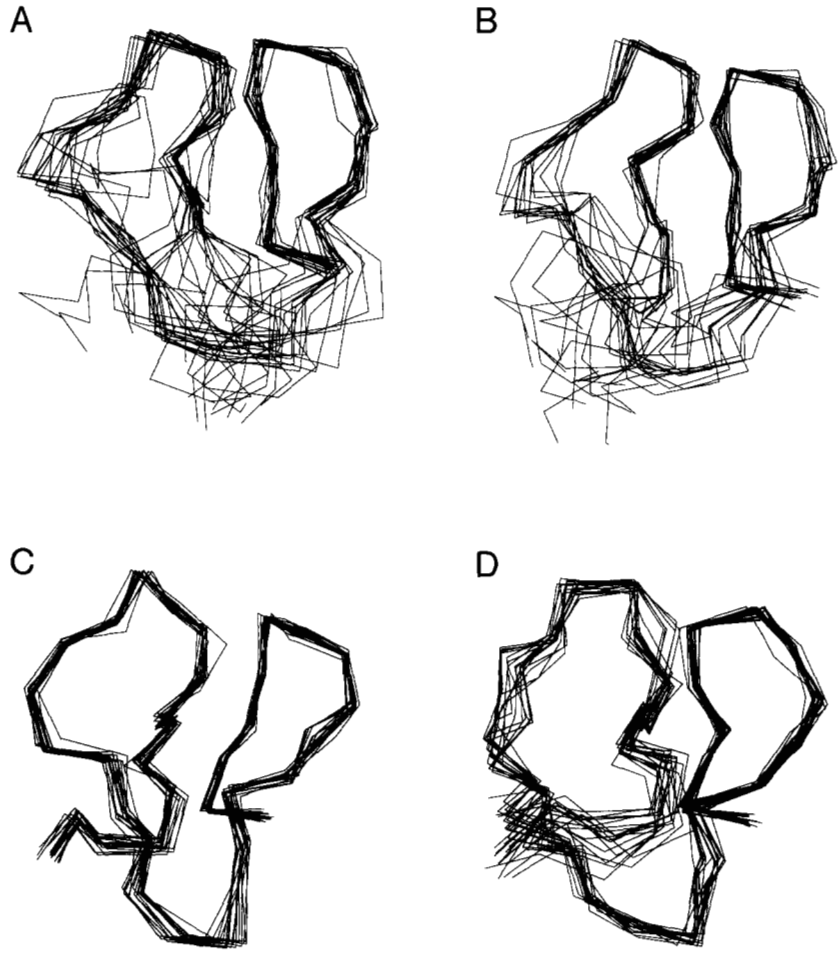

Fig. 3. Structures of Y5A, Y31A, and Y32A. Only backbone C^{α}-traces are shown for clarity. For Y5A the backbone (N, C^{α}, C, O) atoms of residues 7–15 and 25–36 are superimposed. (**A**) 23 and (**B**) 14 structures belonged to two different structural sets. For Y31A (**C**) 18 and for Y32A (**D**) 19 correspondingly superimposed structures are shown.

ertheless, these calculations reproduced the two structural sets for the N-terminus. Twelve structures belonged to the wild-type–like family and the remaining five to the other. A calculation with the ϕ-angle restraint of the wild type also resulted in structures that belonged to the two sets. The structural ambiguity therefore results from the lack of several constraints in the N-terminus, not from the one dihedral constraint only. Whether the N-terminus actually occupies the space on both sides of the β2-strand cannot be settled at present.

In a Ramachandran plot, Y5 ($\phi \sim 60°$, $\psi \sim 20°$) of the wild-type resides outside the regions typically found in proteins (Kraulis et al., 1989). This is caused by a steric strain between the C^{β} atom of Y5 and the carbonyl oxygen of H4. There must, of course, be favorable interactions that compensate for the local unfavorable

Table 1. *1: Average atomic rms deviation of individual structures for the mean structure of backbone and all atoms. 2: Average number of violated distance constraints for interresidue contacts. 3: Average total computed energy*

	Y5A[a]	Y5A[b]	Y31A	Y32A	Wild-type
1. Backbone atoms (Å)	2.2 ± 0.81	2.0 ± 0.83	0.53 ± 0.15	0.81 ± 0.21	0.33 ± 0.04
All atoms (Å)	2.8 ± 0.73	2.5 ± 0.76	1.0 ± 0.13	1.3 ± 0.26	0.52 ± 0.06
2. Inter residue (>0.4 Å)	1 ± 0.6	1 ± 0.6	1 ± 0.5	1 ± 0.5	—
3. Energy (kcal/mol)	60 ± 8	74 ± 12	23 ± 13	24 ± 7	—

The ensemble of the the 22[a] and 14[b] final simulated annealing structures of Y5A. Correspondingly the number of structures were 18 for Y31A and 19 for Y32A. For all engineered peptides structures were refined when disulphide bridges were closed (Cys 8 and Cys 25; Cys 19 and Cys 35). The backbone atoms (N, C^{α}, C, O) of the residues 7–15 and 25–36 were superimposed. Values for the wild-type from Kraulis et al., 1989.

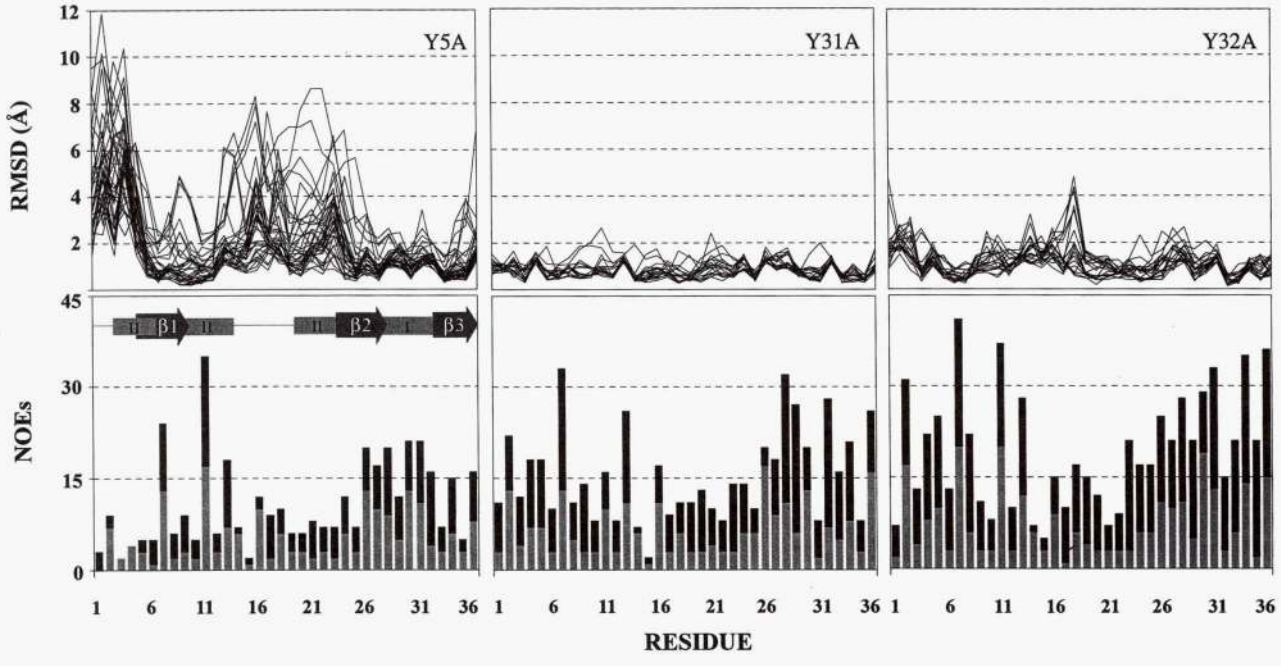

Fig. 4. All atom rms deviations of the individual structures from the average structure, and the corresponding NOEs per residue. Gray columns represent the interresidue NOEs and black columns the intraresidue NOEs. (**A**) 37 (14 + 23) structures of Y5A, (**B**) 18 structures of Y31A, and (**C**) 19 structures of Y32A. The secondary structure elements, type of turns (light gray sticks) and β-strands (dark gray arrows) are indicated at the upper part of the NOE figure of Y5A.

configuration. The aromatic rings of tyrosine 5 and histidine 4 are in parallel orientation, and it is plausible that their interaction stabilizes the unusual ϕ-angle of Y5 (Loewenthal et al., 1992). There are most likely other reasons as well, because H4 is not a conserved amino acid (Linder et al., 1995a). It is of interest to point out that CBD_{EGI} and CBD_{CBHII} have a third disulphide bridge

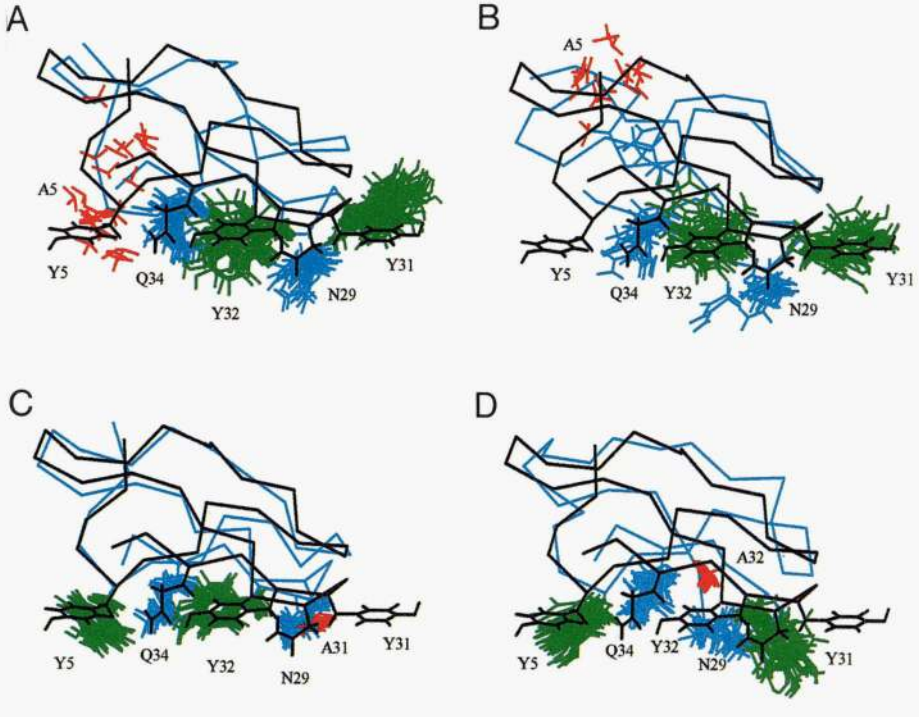

Fig. 5. Side view of the flat face. The backbones of engineered peptides (blue) are superimposed on the wild type (black) and the spatial dispersion of the residues on the flat face are visible. Asparagine (N29) and glutamine (Q34) are indicated in blue. Tyrosines (Y5, Y31, Y32) are shown in green, and the alanines in red. (**A**) 23 and (**B**) 14 structures of Y5A, (**C**) 18 of Y31A, and (**D**) 19 of Y32A.

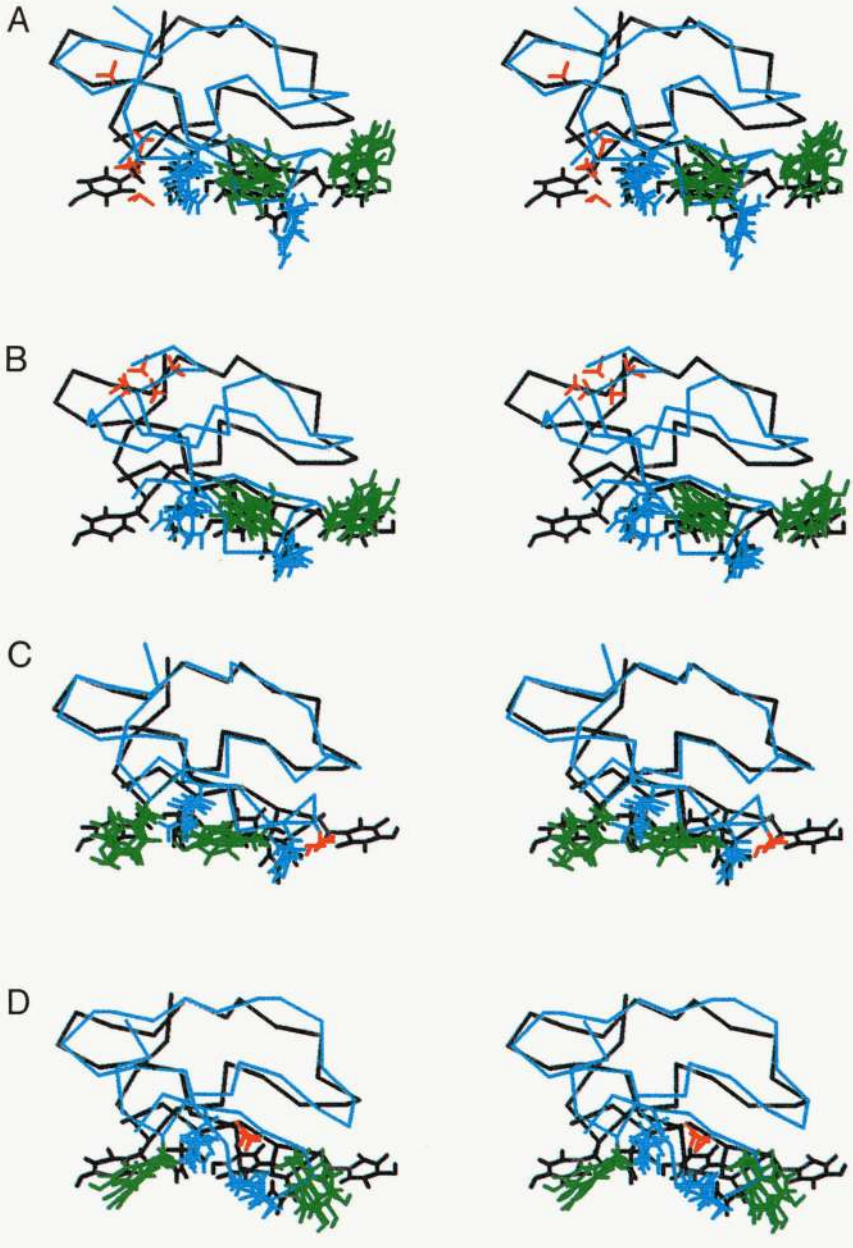

Fig. 6. Stereo side view of the flat face. The backbones of engineered peptides (blue) are superimposed on the wild type (black). For clarity side chains are shown only for five structures. Asparagine (N29) and glutamine (Q34) are indicated in blue and tyrosines (Y5, Y31, Y32) in green. The alanines are shown in red. (**A**) and (**B**) Y5A, (**C**) Y31A, and (**D**) Y32A.

from the tip of the N-terminus to the loop (14–19) region in the middle of the sequence. Although this third disulphide is not conserved, its presence in some of the CBDs may reflect the vulnerability of the N-terminal fold.

For Y31A and Y32A, no ϕ-angle restraint of Y5 could be determined, because no NH—$C^{\alpha}H$ cross-peaks were observed in the COSY spectra. This may well be a result of the cancellation of antiphase lines due to a small spin coupling. Hence, $\phi \approx 60°$ as in the wild-type structure. In the Ramachandran plot, Y5 ($\phi \sim 60°$, $\psi \sim 70°$) of Y31A and ($\phi \sim 100°$, $\psi \sim 80°$) of Y32A are close to the region of Y5 in the wild type even though no ϕ-angle restraints were used. For Y5A the corresponding cross-peak was visible ($^3J_{NH\alpha} = 6.6$ Hz), and the corresponding dihedral angles are $\phi \sim -60°$, $\psi \sim 120°$ and $\phi \sim -100°$, $\psi \sim -170°$ for the best structures in the two conformational families. These angles of A5 are more conventional and different from the backbone dihedrals for Y5 of the wild type. This indicates that primarily for steric reasons, a bulky and rigid aromatic ring at position five is required to maintain the unusual backbone angles and stabile fold in the wild-type structure.

Protein–carbohydrate interactions

The slight difference in pH between the binding studies (Reinikainen et al., 1995) and the structural studies has been shown to be of no importance for correlating structural changes with binding affinities (Linder et al., 1995a). The removal of the aromatic ring

at the tip of the CBD causes no major structural consequences on the remaining residues of the flat face. We suggest that this is the explanation for the small but detectable binding affinity of Y31A to cellulose (Linder et al., 1995a). Because the structural changes are small, the aromatic ring of Y31 has an important contribution to the binding. In the case of Y32A, the local structural changes are inevitable, because removal of the aromatic ring would without side-chain rearrangements leave the interior of CBD exposed. The reorganization of the side chains on the flat face thus causes a complete loss of affinity towards cellulose. The aromatic ring of Y32 undoubtedly contributes to the binding affinity, but the mutation to an alanine also affects other functional residues. In the case of Y5A, it is difficult to partition the role of Y5 into structural or functional. Part of the backbone folding is destroyed and also the remaining structural framework is subject to smaller changes. Consequently, it is not surprising that there is no affinity towards cellulose.

Based on these observations we propose that the binding strength is dictated by the precise positions of the residues on the flat face. The role of this face in the binding of CBD to the surface of cellulose should now be examined by specific substitutions of the tyrosines, for example, by phenylalanines, to reveal the role of hydroxyls, and by histidines and tryptophanes (except W for Y32, which is in the spatially restricted position) to enlighten the role of the hydrophobicity of the aromatic rings. In previous work the binding strength of the Y31H mutant of the CBHI enzyme has been found to be 60% of that for the wild type (Reinikainen et al., 1995). CBD_{EGI} with W5 has roughly twice the affinity for cellulose as CBD_{CBHI} with Y5 (Linder et al., 1995b). Attempts to engineer more strongly binding CBDs could focus on designing larger flat faces. This approach has been exploited by combining two CBDs by a linker peptide (our unpublished results).

In the flat face of type I CBD the periodicity of the aromatic rings and amides and the periodicity of the glucose rings along the polymers of the crystalline cellulose appear complementary (Fig. 7). Four side chains of the wild-type peptide, Y5, Y31, Y32, and Q34, can be aligned well along the cello-oligosaccharide chain. The alignment is within the precision of structure determination, i.e., 0.52 Å (Kraulis et al., 1989). A face-to-face stacking of aromatic rings of amino acids with sugar rings and extensive hydrogen bonding of polar planar groups of amino acid side chains with sugar hydroxyls have been observed in other cases (Quiocho, 1986). The hydrophobic effect has been proposed to account for the improved affinity at increased ionic strength (Reinikainen et al., 1995). Structural motifs for these interactions have also been found, for example, in the hydrolytic sites of the core domains of CBHI (Divine et al., 1994) and CBHII (Rouvinen et al., 1990), as well as in the type II CBD of Cex (Xu et al., 1995).

At the abundant 110 and 1–10 surfaces of crystalline cellulose the sugar polymers stack stepwise on the top of each other so that only parts of the rings are exposed. It is only the obtuse corner of the cellulose crystal that has fully exposed glucose rings (Chanzy et al., 1984; Reinikainen et al., 1995), which would be a prerequisite in the binding model. However, there is experimental evidence that most of the surface of a cellulose crystal is covered by CBD at saturation (Linder et al., 1995b). But it is unclear if CBDs bound on 110 or 1–10 surfaces actually contribute to the crystal breakdown. Face-to-face stacking of aromatic rings to the sugar rings may be required for productive binding. The obtuse corner of the intact cellulose is prone to attack. When the polymer of the obtuse corner becomes solvated, more (020) surface is exposed.

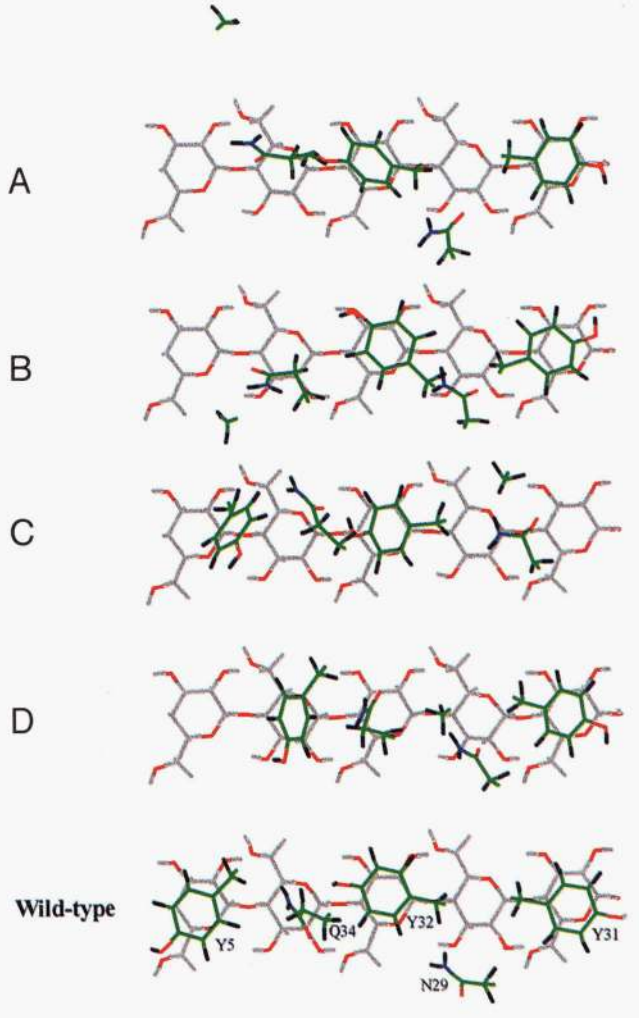

Fig. 7. Engineered peptides (**A–D**) and wild type CBD_{CBHI} placed above the cello-oligosaccharide with aromatic rings stacked on top of sugar rings. Only the residues Y5, N29, Y31, Y32, and Q34 are shown for clarity. The color coding is conventional except in the sugar rings carbons and hydrogens are gray.

Based on the results presently presented, it is concluded that Y31 and Y32 have a functional role for binding to the surface of crystalline cellulose. Their structural role for backbone folding is minor. Y5 has pronounced importance in the structural integrity of the N-terminus of the peptide and probably also functional importance via contributions to the binding affinity. We have thus been able to point out functionally important amino acid side chains, knowledge that can be used in the future for modeling studies of cellulose–cellulase interaction and in designing CBDs with altered binding properties.

Materials and methods

Sample preparation

The synthesis of the three engineered peptides (Y5A, Y31A, and Y32A) and the chemical composition of the purified peptides have been described previously (Linder et al., 1995a). Samples for the NMR experiments were prepared by dissolving 5.0 mg (Y5A), 5.6

mg (Y31A), or 6.5 mg (Y32A) of the peptide in ^1H_2O containing about 10% ^2H_2O for the lock signal. In order to slow down the exchange rate of the amide protons, the pH of each solution was adjusted to 3.9. This same procedure was used in the structure determination of the wild type (Kraulis et al., 1989). NaN_3(0.02%) was added to the samples as an anti-microbial agent. Final sample volume was 0.7 mL.

CD spectroscopy

CD measurements were carried out on a Jasco J-710/720 spectropolarimeter. The samples were diluted from the NMR samples with a 50 μM acetate buffer (pH = 5.0) to obtain peptide concentrations in the range 1–20 μM. Molar ellipticities θ (mdeg) were measured in a 1 mm path length cell as a function of wavelength from 190–250 nm.

^{1}H-NMR spectroscopy

All experiments were performed with a Varian Unity 600 spectrometer. The experiments were conducted at 15 °C, and certain spectra were also acquired at 1 °C and 25 °C. COSY (Marion & Wüthrich, 1983), RELAY-COSY (τ = 35 ms) (Wagner, 1983), NOESY (mixing times: 40, 80, 120, 150 or 160, and 200 ms) (Macura & Ernst, 1980), and sensitivity enhanced TOCSY (mixing times: 30, 50, 80, and 120 ms) (Braunschweiler & Ernst, 1983; Bax & Davis, 1985; Cavanagh & Rance, 1990) were recorded with standard pulse sequences and phase cycling. In addition, NOESY spectra with 300 ms mixing time at 1 °C were measured for Y31A and Y32A. The FIDs were recorded with 2048 complex data points. The number of t_1 increments varied, but usually it was 512 in COSY, 400 in RELAY-COSY, 320 in NOESY, and 256 in TOCSY with 128 or 64 transients per increment. The spectral widths were 6400 Hz in both dimensions. The solvent resonance (4.88 ppm at 15 °C) was suppressed by low-power continuous irradiation during the relaxation delay (1.2 s). The residual signal was further reduced by a time–domain data deconvolution method using a sinebell filter with 20–40 points (Marion et al., 1989; Sodano & Delepierre, 1993). For COSY and RELAY-COSY spectra sinebell weight functions and for NOESY and TOCSY spectra shifted sinebell square weight functions were used in both dimensions. The final data matrices consisted of 2 × 1k real points. The data were Fourier transformed, displayed, and assigned with the Felix data processing and analysis program (Biosym Technologies, San Diego, versions 2.1 and 2.3).

Spectral assignment

The assignment strategy for ^{1}H spin systems of different amino acids was based on the amide proton chemical shift (Chazin et al., 1988). The assignment was partially facilitated by comparisons with the chemical shifts of the wild type (Kraulis et al., 1989), but the chemical shifts of the residues adjacent to the mutations differed too much from the corresponding residues of the wild type to be deduced by comparison. This was also the case for some other residues, especially in the N-termini of the engineered peptides. For these amino acids the assignments were inferred by applying the strategy described by Billeter et al. (1982) and Chazin and Wright (1987). In the case of Y31A, the assignment of the shifts for most protons was obtained from a comparison with the wild type, because residue 31, located at the tip of the molecule, influenced the chemical shifts of the neighboring residues only moderately. For Y5A and Y32A, assignment was somewhat less straightforward, as the ring current effect of the tyrosines affected resonance frequencies of many adjacent protons. The complete sequence and stereo-specific assignments were obtained for all mutants. The wild-type–like pairing of the two disulphide bridges was also confirmed for the engineered peptides from the NOEs of $C^{\beta}H$ of the cysteines.

NMR-derived constraints

For every unambiguously assigned and well-resolved cross-peak the initial slope of the NOE build-up curve was calculated from a fitted second-order polynomial and converted into distances by calibrating the initial slope to distance. Covalent distances between certain geminal protons were used for the calibration (1.78 Å). The determined distances were given with ±20% uncertainty ranges. Intensities of protons with degenerate shifts were divided by the number of equivalent protons (e.g., by three in the case of a methyl group) before the distances were extracted and an extra 0.5 Å was added to obtain the upper bounds (Clore et al., 1987; Wagner et al., 1987). Distances corresponding to weak or poorly resolved cross-peaks were given lower bounds of at 1.8 Å, and upper bounds 6.0 Å. For Y31A and Y32A, more cross-peaks were identified from the NOESY spectra recorded at 1 °C and with a 300 ms mixing time. The restraint sets were supplemented with these qualitative restraints with lower and upper bounds of 1.8 Å and 5.0 Å, respectively, and extended by 1 Å for pseudo-atoms.

$^3J_{HN\alpha}$and $^3J_{\alpha\beta}$ coupling constants were measured from COSY spectra in ^1H_2O by the *J*-doubling method (Le Parco et al., 1992; McIntyre & Freeman, 1992). For Y32A, $^3J_{\alpha\beta}$ coupling constants had to be determined from the spectrum recorded in ^2H_2O, as the $C^{\alpha}H$-$C^{\beta}H$ region was disturbed by the residual water signal. Stereospecific assignments of the β-methylene protons were obtained based on NOEs if one of the $^3J_{\alpha\beta}$ was greater than 10 Hz or both were smaller than 5 Hz (Wagner et al., 1987). The $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ coupling constants were converted to ϕ- and χ-angles by means of Karplus relations (deMarco et al., 1978; Smith et al., 1996). The final dihedral angle restraints were obtained by adding a ±30° margin to the determined dihedral angle.

Structure calculations

The number of distance restraints imposed during calculations and the corresponding values for the wild type (Kraulis et al., 1989) are summarized in Table 2. The total number of distance restraints was 524 for Y31A and 519 for Y32A. For Y5A only 295 distance restraints could be identified. This mutant also had significantly fewer dihedral angle restraints (32) than Y31A (42), Y32A (42), and the wild type (82). The final set of structures was computed with a total of 327 experimental restraints for Y5A, 566 for Y31A, and 561 for Y32A. The corresponding value for the wild type was 660, which included 24 hydrogen bond restraints imposed after structure generation assuming certain acceptor oxygens, and 24 Ψ-angle restraints obtained by modeling. These were not used in our structure determinations. Therefore, in our calculations the number of experimental restraints for Y31A and Y32A were comparable to wild type. For Y5A, the number of experimental restraints was only about 60% of the restraints for the other peptides.

Initially ensembles of 60 (Y5A), 40 (Y31A), and 40 (Y32A) structures were generated by the standard protocols of the distance

Table 2. *Number of experimental restraints used for structure calculations*

		Distance			Dihedral				
Peptide	ISPA	Intensity	Qualitative	Total	ϕ	χ	ψ	H-bond	Total
Y5A	295		0	295	22	10	0	0	327
Intraresidue	196								
Interresidue, short range ($\|i-j\|$[a]) ≤ 5	75								
Interresidue, long range ($\|i-j\|$) > 5	24								
Y31A	397		127	524	24	18	0	0	566
Intraresidue	237								
Interresidue, short range ($\|i-j\|$) ≤ 5	116								
Interresidue, long range ($\|i-j\|$) > 5	44								
Y32A	494		25	519	24	18	0	0	561
Intraresidue	279								
Interresidue, short range ($\|i-j\|$) ≤ 5	138								
Interresidue, long range ($\|i-j\|$) > 5	77								
Wild-type[b]		554	0	554	33	25	24	24	660
Intraresidue		211							
Interresidue, short range ($\|i-j\|$) ≤ 5		206							
Interresidue, long range ($\|i-j\|$) > 5		137							

[a]Residues i and j belong to different spin systems.
[b]Values obtained from Kraulis et al., 1989.

geometry (DGII, Discover 3.1, Biosym Technologies, San Diego, CA) (Crippen, 1977; Havel et al., 1983, 1990, 1991; Sheeck et al., 1991). Because Y5A had fewer restraints than the other engineered peptides, it was necessary to map the conformation space with a larger number of structures. The stereospecific assignments of glycine α protons were deduced after the initial structure generation and later imposed. Upon inspection of the initial structures new cross-peaks were identified. The DGII calculations were then repeated and the structures refined by simulated annealing (SA, Discover 3.1, Biosym Technologies, San Diego, CA) (Clore et al., 1986; Nilges et al., 1988a, 1988b, 1988c). In the molecular dynamics (MD) computation, the AMBER force field was used. SA/MD structures were finally minimized by a conjugate gradients algorithm. The structures were inspected with the InsightII (Biosym Technologies, San Diego, CA) molecular modeling software to detect inconsistencies and restraint violations. All calculations and graphical analyses were performed on Silicon Graphics IRIS 4D30 TG and Indigo 2 workstations.

Acknowledgments

The co-ordinates for Y5A, Y31A, and Y32A will be deposited in the Brookhaven Protein Data Bank and the chemical shifts in the BioMagResBank at Madison. This work has been supported by the Academy of Finland, the Emil Aaltonen Foundation, the foundation of Ella and Georg Ehrnrooth, the Swedish National Board for Industrial and Technical Development, and the Neste Research Foundation.

References

Bax A, Davis DG. 1985. MLEV-17 based two-dimensional homonuclear magnetisation transfer spectroscopy. *J Magn Reson 65*:355–360.

Béguin P, Aubert PJ. 1994. The biological degradation of cellulose. *FEMS Microbiol Rev 13*:25–58.

Billeter M, Braun W, Wüthrich K. 1982. Sequential resonance assignments in protein proton nuclear magnetic resonance spectra. Computation of sterically allowed proton-proton distances and statistical analysis of proton-proton distances in single crystal protein conformations. *J Mol Biol 155*:321–346.

Braunschweiler L, Ernst RR. 1983. Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J Magn Reson 53*:521–528.

Cambillau C, van Tilbeurgh H. 1993. Structure of hydrolases: Lipases and cellulases. *Curr Opin Struct Biol 3*:885–895.

Cavanagh J, Rance M. 1990. Sensitivity improvement in isotropic mixing (TOCSY) experiments. *J Magn Reson 88*:72–85.

Chanzy H, Henrissat B, Vuong R, Schülein M. 1984. The action of 1,4-β-D-glucan cellobiohydrolase on *Valonia* cellulose microcrystals. An electron microscopic study. *FEBS Lett 153*:113–118.

Chazin WJ, Wright PE. 1987. A modified strategy for identification of proton spin systems in proteins. *Biopolymers 26*:973–977.

Chazin WJ, Rance M, Wright PE. 1988. Complete assignment of the ^{1}H nuclear magnetic resonance spectrum of french bean plastocyanin. Application of an integrated approach to spin system identification in proteins. *J Mol Biol 202*:603–622.

Claeyssens M, Tomme P. 1989. Structure–function relationships of cellulolytic proteins from *Trichoderma reesei*. In: Kubicek CP, ed. *Trichoderma reesei cellulases: Biochemistry, genetics, physiology and application*. Cambridge: Royal Society of Chemistry. pp 1–11.

Clore GM, Brünger AT, Karplus M, Gronenborn AM. 1986. Application of molecular dynamics with interproton distance restraints to three-dimensional protein structure determination: A model study of crambin. *J Mol Biol 191*:523–551.

Clore GM, Gronenborn AM, Nilges M, Ryan CA. 1987. The three-dimensional structure of potato carboxypeptidase inhibitor in solution: A study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *Biochemistry 26*:8012–8023.

Crippen GM. 1977. A novel approach to the calculation of conformation: Distance geometry. *J Comp Phys 24*:96–107.

deMarco A, Linas M, Wüthrich K. 1978. Analysis of the ^{1}H-NMR spectra of ferrichrome peptides. I. The non-amide protons. *Biopolymers 17*:617–636.

Din N, Forsythe IJ, Burtnick LD, Gilkes NR, Miller RC, Warren RAJ, Kilburn DG. 1994. The cellulose-binding domain of endoglucanase A (CenA) from *Cellumonas fimi*: Evidence for involvement of tryptophan residues in binding. *Mol Microbiol 11*:747–755.

Divine C, Ståhlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JKC, Teeri TT, Jones TA. 1994. The three-dimensional structure of cellobiohydrolase I from *Trichoderma reesei* reveals a lysozyme-like active site on a lecitin-like framework. *Science 265*:524–528.

Gardner KH, Blackwell J. 1974. The structure of native cellulose. *Biopolymers 13*:1975–2001.

Havel TF. 1990. The sampling properties of some distance geometry algorithms

applied to unconstrained polypeptide chains: A study of 1830 independently computed conformations. *Biopolymers 29*:1565–1585.

Havel TF. 1991. An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. *Prog Biophys Mol Biol 56*:43–78.

Havel TF, Kuntz ID, Crippen GM. 1983. The theory and practice of distance geometry. *Bull Math Biol 45*:665–720.

Henrissat B, Driguez H, Viet C, Schülein M. 1985. Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. *Bio/Technology 3*:722–726.

Irwin DC, Spezio M, Walker LP, Wilson DB. 1993. Activity studies of eight purified cellulases: Specificity, synergism and binding domain efforts. *Biotechnol Bioeng 42*:1002–1013.

Kraulis PJ, Clore MG, Nilges M, Jones AT, Pettersson G, Knowles J, Gronenborn AM. 1989. Determination of the three-dimensional structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. *Biochemistry 28*:7241–7257.

Le Parco JM, McIntyre L, Freeman R. 1992. Accurate coupling constants from two-dimensional correlation spectra by "I deconvulsion." *J Magn Reson 97*:553–567.

Linder M, Mattinen ML, Kontteli M, Lindeberg G, Ståhlberg J, Drakenberg T, Reinikainen T, Pettersson G, Annila A. 1995a. Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. *Protein Sci 4*:1056–1064.

Linder M, Lindeberg G, Reinikainen T, Teeri T, Pettersson G. 1995b. The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution. *FEBS Lett 372*:96–98.

Loewenthal R, Sancho J, Fersht AR. 1992. Histidine-aromatic interactions in barnase. *J Mol Biol 224*:759–770.

Macura A, Ernst RR. 1980. Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Mol Phys 41*:95–117.

Marion D, Wüthrich K. 1983. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^1H-^1H spin-spin coupling constants. *Biochem Biophys Res Commun 113*:967–974.

Marion D, Mitsuhiko I, Bax A. 1989. Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data. *J Magn Reson 84*:425–430.

McIntyre L, Freeman R. 1992. Accurate measurement of coupling constants by J doubling. *J Magn Reson 96*:425–431.

Nidetzky B, Hayn M, Macarron R, Steiner W. 1993. Synergism of *Trichoderma reesei* cellulases while degrading different celluloses. *Biotechnol Lett 15*:71–76.

Nilges M, Clore GM, Gronenborn AM. 1988a. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett 229*:317–324.

Nilges M, Gronenborn AM, Brünger AT, Clore GM. 1988b. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng 2*:27–38.

Nilges M, Clore GM, Gronenborn AM. 1988c. Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. *FEBS Lett 239*:129–136.

Poole DM, Hazlewood GP, Huskisson NS, Virden R, Gilbert HJ. 1993. The role of conserved tryptophan residue in the interaction of a bacterial cellulose binding domain with its ligand. *FEMS Microbiol Lett 106*:77–84.

Quiocho FA. 1986. Carbohydrate-binding proteins: Tertiary structures and protein–sugar interactions. *Annu Rev Biochem 55*:287–315.

Quiocho FA. 1993. Probing the atomic interactions between proteins and carbohydrates. *Biochem Soc Transact 21*:442–448.

Reinikainen T, Ruohonen L, Nevanen T, Laaksonen L, Kraulis P, Jones TA, Knowles JKC, Teeri T. 1992. Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase I. *Proteins Struct Funct Genet 14*:475–482.

Reinikainen T, Teleman O, Teeri TT. 1995. Effects of pH and high ionic strength on the adsorption and activity of native and mutated cellobiohydrolase I from *Trichoderma reesei*. *Proteins 22*:392–403.

Rouvinen J, Bergfors T, Teeri T, Knowles JKC, Jones TA. 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science 249*:380–386.

Sheeck RM, Torda AE, Kemmink J, van Gunsteren WF. 1991. Structure determination by NMR: The modelling of NMR parameters as ensemble averages. In: Hoch JC, Poulsen FM, Redfield C, eds. *Computational aspects of the study of biological macromolecules by nuclear magnetic resonance spectroscopy*. New York: Plenum Press. pp 209–217.

Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM. 1996. Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. *J Mol Biol 255*:494–506.

Sodano P, Delepierre M. 1993. Clean and efficient suppression of the water signal in multidimensional NMR spectra. *J Magn Reson 104*:88–92.

Ståhlberg J, Johansson G, Pettersson G. 1988. A binding-site-deficient, catalytically active, core protein of endoglucanase III from the culture filtrate of *Trichoderma reesei*. *Eur J Biochem 173*:179–183.

Ståhlberg J, Johansson G, Pettersson G. 1991. A new model for enzymatic hydrolysis of cellulose based on the two-domain structure of cellobiohydrolase I. *Bio/Technology 9*:286–290.

Tomme P, Tilbeurgh H van, Pettersson G, Damme J van, Vandekerckhove J, Knowles J, Teeri T, Claeyssens M. 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur J Biochem 170*:575–581.

van Tilbeurgh H, Tomme P, Claeyssens M, Bhikhabhai R, Pettersson G. 1986. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. *FEBS Lett 204*:223–227.

Vyas NK. 1991. Atomic features of protein–carbohydrate interactions. *Curr Opin Struct Biol 1*:723–740.

Wagner G. 1983. Two-dimensional relayed coherence transfer spectroscopy of a protein. *J Magn Reson 55*:151–156.

Wagner G, Braun W, Havel TF, Schaumann T, Go N, Wüthrich K. 1987. Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J Mol Biol 196*:611–639.

Wood TM, Garcia-Campayo V. 1990. Enzymology of cellulose degradation. *Biodegradation 1*:147–161.

Xu GY, Ong E, Gilkes NR, Kilburn DG, Muhandiram DR, Harris-Brandts M, Carver JP, Kay LE, Havery TS. 1995. Solution structure of a cellulose-binding domain from *Cellumonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry 34*:6993–7009.