Clostridium thermocellum Xyn10B Carbohydrate-Binding Module 22-2: The Role of Conserved Amino Acids in Ligand Binding † , ‡

ARTICLE *in* BIOCHEMISTRY · SEPTEMBER 2001

Impact Factor: 3.02 · DOI: 10.1021/bi0106742 · Source: PubMed

Clostridium thermocellum Xyn10B Carbohydrate-Binding Module 22-2: The Role of Conserved Amino Acids in Ligand Binding[†,‡]

Hefang Xie,[§] Harry J. Gilbert,[§] Simon J. Charnock,[||,†] Gideon J. Davies,[||] Michael P. Williamson,[⊥] Peter J. Simpson,[⊥] S. Raghothama,[⊥] Carlos M. G. A. Fontes,[▽] Fernando M. V. Dias,[▽] Luis M. A. Ferreira,[▽] and David N. Bolam*[,§]

Department of Biological and Nutritional Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, U.K., Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, U.K., Department of Molecular Biology and Biotechnology, Krebs Institute, University of Sheffield, Sheffield S10 2TN, U.K., and CIISA-Faculdade de Medicina Veterinária, Rua Prof. Cid dos Santos, 1300-477 Lisboa, Portugal

Received April 3, 2001; Revised Manuscript Received May 24, 2001

ABSTRACT: The majority of plant cell wall hydrolases are modular enzymes which, in addition to a catalytic module, possess one or more carbohydrate-binding modules (CBMs). These carbohydrate-active enzymes and their constituent modules have been classified into a number of families based upon amino acid sequence similarity. The *Clostridium thermocellum* xylanase, Xyn10B, contains two CBMs that belong to family 22 (CBM22). The crystal structure of the C-terminal CBM22 (CBM22-2) was determined in a previous study [Charnock, S. J., et al. (2000) *Biochemistry 39*, 5013−5021] and revealed a surface cleft which presents several conserved residues that are implicated in ligand binding. These amino acids have been substituted and the structure and biochemical properties of the mutants analyzed. The data show that R25A, W53A, Y103A, Y136A, and E138A exhibit greatly reduced affinity for xylotetraose relative to that of the wild-type protein. Conversely, mutations Y103F and Y136F have little effect on ligand binding. Using thermodynamic, X-ray, and NMR measurements on the mutants, we show that the cleft of CBM22-2 does indeed form the ligand-binding site. Trp 53 and Tyr 103 most likely participate in hydrophobic stacking interactions with the ligand, while Glu 138 makes one or more important hydrogen bonds with the tetrasaccharide. Although Arg 25 and Tyr 136 are likely to form hydrogen bonds with the ligand, they are also shown to play a critical role in maintaining the structural integrity of the binding cleft.

The plant cell wall is a composite insoluble structure that is highly recalcitrant to biological degradation. Microbial enzymes involved in the recycling of photosynthetically fixed carbon have evolved specific features which enable them to digest this highly recalcitrant polymer. One of the main features of these enzymes that distinguishes them from most other biocatalysts is their modular architecture (*1*). They generally contain a combination of catalytic and noncatalytic modules in a single polypeptide chain, all of which contribute to the efficiency of enzyme catalysis. The most common noncatalytic modules in plant cell wall hydrolases are carbohydrate-binding modules (CBMs; *1*). CBMs have been classified into 25 distinct families based on primary sequence conservation (*2*). Furthermore, there are a vast number of uncharacterized modules, termed "X" modules, whose function may well involve carbohydrate binding (*3*).

A large number of characterized CBMs bind to crystalline cellulose, exemplified by members of families 1, 2a, 3a, 5, and 10; however, recent studies have shown that CBMs appended to xylanase catalytic modules often bind preferentially to xylan (*4*−*11*). CBM families whose members bind xylan include CBM2b, CBM4, CBM6, CBM13, and CBM22. Irrespective of ligand specificity, it appears that most CBMs mediate increased catalytic efficiency by increasing the concentration of the enzyme on the surface of insoluble substrates (*11*−*13*). Some CBMs may also display additional functions such as substrate disruption or the sequestering and feeding of single glycan chains into the active site of the adjacent catalytic module (*14*−*17*).

The three-dimensional structures of numerous CBMs have so far demonstrated that these modules are composed exclusively of β-strands. Modules that bind to crystalline cellulose contain a planar hydrophobic platform comprising aromatic residues that bind to one of the surfaces of the

[†] This work was funded, in part, by the Biotechnology and Biological Sciences Research Council. G.J.D. is a Royal Society University Research Fellow.

[‡] Coordinates for the structures described in this paper have been deposited with the macromolecular structures database (entries 1h6x and 1h6y).

* To whom correspondence should be addressed. Telephone: +44-191-2226962. Fax: +44-191-2228684. E-mail: D.N.Bolam@ncl.ac.uk.

[§] University of Newcastle upon Tyne.

[||] University of York.

[⊥] University of Sheffield.

[▽] CIISA-Faculdade de Medicina Veterinária.

[1] Abbreviations: AGE, affinity gel electrophoresis; CBM, carbohydrate-binding module; CM, catalytic module; IMAC, immobilized metal affinity chromatography; ITC, isothermal titration calorimetry; NMR, nuclear magnetic resonance; PNPC, 4-nitrophenyl β-D-cellobioside.

10.1021/bi0106742 CCC: $20.00
Published on Web 07/12/2001

polysaccharide (*4*, *5*, *18*, *19*). CBMs that bind individual polymer chains usually possess clefts which either have been shown, or are presumed, to be the ligand binding site (*11*, *20*, *21*). In common with many other carbohydrate-active proteins, aromatic residues play a pivotal role in ligand binding (*6*, *19*, *22*−*26*). Indeed, changes in the orientation of these aromatic groups can determine their specificity for cellulose or xylan (*7*). In contrast to the importance of aromatic residues in CBM−carbohydrate interactions, the role of polar amino acids is less clear. Several studies have shown that the mutation of amino acids which may participate in hydrogen bonding to the target ligand (located on the surface binding site) often results in only a modest reduction in affinity (*19*, *25*, *26*).

Recently, it has been demonstrated that the X6 family of proteins, previously of unknown function, contains members that bind xylan and xylooligosaccharides (*11*, *27*, *28*). Thus, X6 modules were reclassified as family 22 CBMs (*11*). The xylanase Xyn10B (formally XynY) from *Clostridium thermocellum* contains two CBM22s (CBM22-1 and CBM22-2, formerly X6a and X6b, respectively) flanking a GH10 catalytic module, a dockerin sequence, and a C-terminal carbohydrate/ferulate esterase catalytic module (CE1; *29*). CBM22-2, which is located C-terminal of GH10CM, interacts with a range of xylans and xylooligosaccharides. The crystal structure of CBM22-2 revealed a classic β-jelly roll fold with a cleft running along its entire length that contains three aromatic and two polar residues which are conserved in many other CBM22 proteins. We had previously suggested, on the basis of analogy with other glycan-chain binding modules, that this cleft comprises the ligand binding site and that amino acids located within the cleft were therefore likely to be critical for carbohydrate recognition (*11*). Here we describe a site-directed mutagenesis approach to investigating the importance of these conserved residues in ligand binding. The data show that Trp 53, Tyr 103, and Glu 138 play pivotal roles in the interaction of CBM22-2 with xylan, while Tyr 136 and Arg 25, although involved in ligand binding, also maintain the structural integrity of the cleft.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Culture Conditions. The *Escherichia coli* strain used in this study was BL21(DE3): pLysS. *E. coli* strains were cultured in Luria broth supplemented with 100 μg/mL ampicillin at 37 °C with shaking at 180 rpm. Expression of Xyn10B derivatives using isopropyl β-thiogalactopyranoside was as described previously (*11*). The *E. coli* strains used in these expression studies harbored recombinants of the pET21a expression vector (Novagen) containing truncated forms of the Xyn10B gene (*xyn10B*). The construction of recombinant plasmids pCF1, pCF2, pCF3, and pCF5, encoding CBM22-1-GH10CM, GH10CM, GH10CM-CBM22-2, and CBM22-2, respectively, was described previously (*11*). The order of the modules in each of the constructs is the same as that which occurs in the full-length enzyme. The pCF6 plasmid, encoding CBM22-1, is described below.

Production of pCF6 and Protein Purification. To express CBM22-1, the region of *xyn10B* encoding this module was amplified by PCR using the reaction conditions described previously (*11*) and the following primers which contain *Nhe*I and *Xho*I restriction sites, respectively: 5′-CTCGCTAGCGATTATGAAGTGGTTCATG-3′ and 5′-CACCTCGAGATACATATCTTTCAGCACAG-3′. The amplified DNA was cloned into *Nhe*I- and *Xho*I-digested pET21a.

The Quick Change mutagenesis kit (Stratagene) was used to generate the mutants of CBM22-2 using pCF5 as the template DNA. The primer pairs used to produce the required mutations were as follows: 5′-CGTAGGACAGTGGACTGCA**GCA**GGACCTGCGGAAG-3′ and 5′-CTTCCGCAGGTCC**TG**CTGCAGTCCACTGTCCTACG-3′ (*Pst*I) for R25A, 5′-GTAAGGAACCGTACTGCAGCA**GCG**AACGGAGCACAACG-3′ and 5′-CGTTGTGCTCCGTT**CGC**TGCTGCAGTACGGTTCCTTAC-3′ (*Pst*I) for W53A, 5′-GGCACTCAACGG**GCT**GATACGATCGATATGAAAACTG-3′ and 5′-CGATTTTCATATCGATCGTATC**AGC**CCGTTGAGTGCC-3′ (*Pvu*I) for Y103A, 5′-GCAACAGATATG**GCT**GTTTACGTAGAAACAGCGGATG-3′ and 5′-CATCCGCTGTTTCTACGTAAAC**AGC**CATATCTGTTGC-3′ (*Sna*BI) for Y134A, 5′-GCAACAGATATGTACGTA**GCT**GTGGAAACAGCG-3′ and 5′-CGCTGTTTCCAC**AGC**TACGTACATATCTGTTGC-3′ (*Sna*BI) for Y136A, 5′-GTATGTTTATGTG**GCG**ACTGCAGATGACACCATTAAC-3′ and 5′-GTTAATGGTGTCATCTGCAGT**CGC**CACATAAACATAC-3′ (*Pst*I) for E138A, 5′-GGCACTCAACGG**TTT**GATACGATCGATATGAAAACTG-3′ and 5′-CAGTTTTCATATCGATCGTATC**AAA**CCGTTGAGTGCC-3′ (*Pvu*I) for Y103F, and 5′-GCAACAGATATGTACGTA**TTT**GTGGAAACAGCG-3′ and 5′-CGCTGTTTCCAC**AAA**TACGTACATATCTGTTGC-3′ (*Sna*BI) for Y136F. The mutated nucleotides are in bold. The primers contain restriction sites (underlined), which were used in the initial selection of the mutants. All derivatives of Xyn10B were purified by immobilized metal affinity chromatography (IMAC). Briefly, cell-free extract (20 mL) was passed through a 5 mL bed volume Talon (ClonTech) column and washed with 50 mL of 20 mM Tris-HCl (pH 8.0) containing 100 mM NaCl (buffer A) and the purified His-tagged CBM eluted with 15 mL of 100 mM imidazole in buffer A. The proteins were then dialyzed against 50 mM sodium phosphate buffer (pH 7.0) and stored at 4 °C.

Protease and Thermostability Assays. Approximately 600 ng of purified CBM22-1-GH10CM and GH10CM-CBM22-2 and 400 ng of purified GH10CM were incubated with porcine pancreatin (Sigma) at a final concentration of 10 mg/mL in 50 mM sodium phosphate buffer (pH 7.0) at 37 °C. At regular time intervals, an aliquot of the reaction mixture was removed and assayed for enzyme activity at 50 °C by measuring the amount of reducing sugars released from soluble oat spelt xylan as described by Miller (*30*). For the thermostability experiments, the Xyn10B derivatives were incubated at 65 °C in 50 mM sodium phosphate buffer (pH 7.0), and at regular intervals, residual enzyme activity was evaluated as for protease resistance assays.

Identification of Sugars Released from Xylan by Xyn10B Derivatives. HPLC was carried out on a Dionex system using a CarboPac PA-100 column at a flow rate of 1 mL/min. Enzyme reactions were carried out as follows. CBM22-1-GH10CM and GH10CM were first matched in activity against 10 mM *p*-nitrophenyl β-D-cellobioside (PNPC; Sigma) in 50 mM sodium phosphate buffer (pH 7.0) containing 1 mg/mL bovine serum albumin (BSA; Sigma).

Matched amounts of enzyme (0.3−0.5 μM) were incubated at 37 °C with either 2% (w/v) soluble oat spelt xylan, 2% (w/v) insoluble unsubstituted xylan, or 0.1 mM xylooligosaccharides in 50 mM sodium phosphate buffer (pH 7.0) containing 1 mg/mL BSA. Aliquots were removed at regular time intervals and boiled for 10 min to inactivate the enzyme. Samples were then subjected to HPLC. Separation of xylooligosaccharides was achieved using a gradient of 0 to 75 mM sodium acetate in 100 mM sodium hydroxide from 0 to 15 min, followed by isocratic flow with 75 mM sodium acetate (in 100 mM sodium hydroxide) from 16 to 25 min. Sugars were detected by pulsed amperometry and identified by comparison with the elution profile of standard xylooligosaccharides.

Protein Quantification and Assessment of Purity. Protein concentrations were determined from their molar extinction coefficients at 280 nm (*31*), which were as follows: CBM22-1, 24 400 M^{-1} cm^{-1}; CBM22-1-GH10CM, 96 000 M^{-1} cm^{-1}; GH10CM, 71 600 M^{-1} cm^{-1}; GH10CM-CBM22-2, 104 300 M^{-1} cm^{-1}; and CBM22-2, 32 700 M^{-1} cm^{-1}. Protein purity was determined by SDS−PAGE.

Sources of Sugars. Xylooligosaccharides were from Megazyme. The soluble fraction of oat spelt xylan (Sigma) was prepared as described previously (*32*). Insoluble unsubstituted xylan was prepared as follows. To 8 mL of 2% (w/v) oat spelt xylan (approximately 10% arabinose) in 50 mM sodium phosphate buffer (pH 7.0) was added 200 units of a xylan-specific arabinofuranosidase (*Pseudomonas cellulosa* Abf62A). Aliquots were removed at regular time intervals and subjected to HPLC (as described above) to determine when all the accessible arabinose side chains had been removed. After removal of the arabinose side chains, the xylan became insoluble. The insoluble unsubstituted xylan was washed twice in 10 mL of 50 mM sodium phosphate buffer (pH 7.0) and finally resuspended in 4 mL of the same buffer.

Ligand Binding Studies. Affinity gel electrophoresis (AGE) was used to qualitatively assess the capacity of CMB22-2 mutants to bind xylan. AGE was performed as described previously (*11*). Soluble oat spelt xylan at a final concentration of 0.1% was used as the ligand. Isothermal titration calorimetry (ITC) was performed as described previously using xylotetraose as the ligand with 50−100 μM protein (*21*).

NMR Spectroscopy. Protein samples were approximately 700 μM in 10 mM sodium phosphate buffer (pH 7.0) containing 10% D_2O. NMR spectra were recorded at 30 °C on a Bruker DRX-500 spectrometer, and ^1H chemical shifts were referenced to an internal standard of 3-(trimethylsilyl)-2,2,3,3-d_4-propionate at 0.00 ppm. Data were processed by FELIX (MSI). Binding affinity was measured by titrating xylotetraose into the protein and fitting the chemical shift changes to a one-site binding model using Microsoft Excel version 5.0.

Crystallization and Data Collection. Crystals of the *C. thermocellum* Xyn10B CBM22-2 mutants R25A and E138A were grown by vapor diffusion using the hanging drop method. R25A crystallized after 2 days from a 2 μL drop comprising equal volumes of protein at 50 mg/mL and 1.7 M ammonium sulfate. E138A was crystallized essentially as described previously for seleno-L-methionyl CBM22-2 (*11*) after microseeding with native crystals. A cryogenic storage solution for R25A comprised the growth conditions supplemented with 25% (v/v) glycerol, in which the crystals were soaked for 10 s prior to mounting in a rayon fiber loop and freezing in liquid nitrogen. Crystals of E138A were frozen directly from their mother liquor. Preliminary X-ray diffraction analysis conducted on a home source revealed that R25A crystallized into a tetragonal space group, with the following unit cell dimensions: $a = b = 38.68$ Å and $c = 207.53$ Å; E138A crystallized in space group $P6_122$, as for the wild-type enzyme, albeit with the following slightly changed unit cell dimensions: $a = b = 89.87$ Å and $c = 209.00$ Å (native cell, $a = b = 90.08$ Å and $c = 207.71$ Å). Data to 2.25 and 2.20 Å, for R25A and E138A, respectively, were collected on beamline ID14-1 at the European Synchrotron Radiation Facility (Grenoble, France) using an ADSC CCD detector.

Data Processing and Structure Determination. Data were processed using the HKL2000 program suite (*33*). All further computing was carried out using the CCP4 suite of programs (*34*). The structure of R25A was determined by molecular replacement using the program AMoRe (*35*) with the native CBM22-2 used as the search model. Structures were refined using REFMAC (*36*), having first set aside 5% of the observations for cross-validation analysis. Water molecules were added automatically and verified manually prior to coordinate deposition. The structure of E138A was determined by directly refining the published native structure of CBM22-2 with the mutant data, maintaining the same cross-validation subset of reflections as used for the native structure determination.

Phylogenetic Analysis. Construction of the phylogenetic tree and phylogenetic analysis were performed by alignment of the available CBM22 homologues with ClustalW (*37*) followed by tree construction with PUZZLE (*38*), using quartet sampling and neighbor options.

RESULTS

Proposed Binding Cleft of CBM22-2. In a previous study, we showed that CBM22-2 from *C. thermocellum* Xyn10B bound to soluble xylan and a range of xylooligosaccharides (*11*). The structure revealed a cleft that was proposed to be the xylan binding site. Inspection of the surface of the cleft revealed several amino acids which were conserved in other CBM22 members (Figure 1). These amino acids include three residues, Arg 25, Glu 138, and Tyr 136 (whose aromatic ring is buried but whose hydroxyl group is fully exposed on the surface), which have the potential to form hydrogen bonds with the target ligand, while Trp 53 and Tyr 103 could interact via hydrophobic stacking interactions. A fourth aromatic residue, Tyr 134, is also on the surface of the cleft but is less highly conserved than the other amino acids described above.

To evaluate the importance of the amino acids on the surface of the CBM22-2 cleft in ligand binding, mutants R25A, W53A, Y103A, Y103F, Y134A, Y136A, Y136F, and E138A were constructed, and their biochemical properties studied. The eight mutants were expressed efficiently in *E. coli* and were purified using IMAC to apparent homogeneity as judged by SDS−PAGE (data not shown).

Structural Integrity of the Mutants. The structural integrity of the mutant proteins was determined by NMR (Figure 2A). The upfield ends of the different spectra, ranging from −0.54

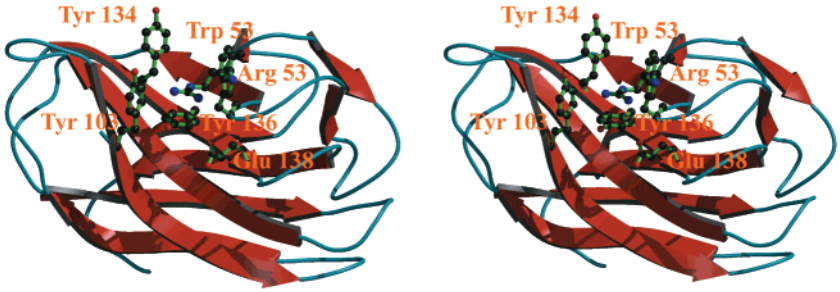

FIGURE 1: Structure of CBM22-2. Divergent (wall-eyed) stereo cartoon of the CBM22-2 structure. Residues mutated in this study are shown in ball-and-stick representation. The figure was drawn with BOBSCRIPT (*45*) and rendered with RASTER 3-D (*46*).

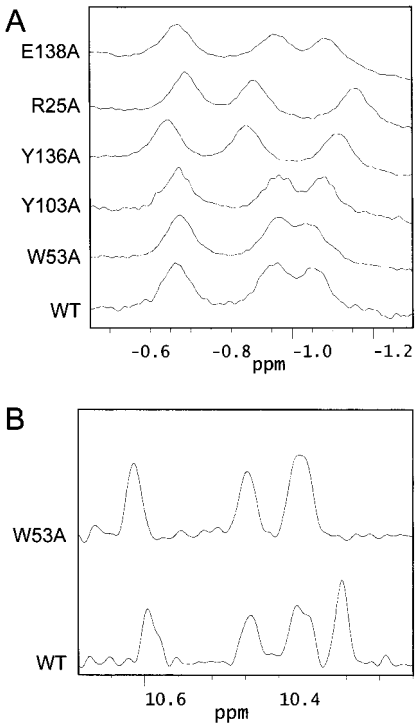

FIGURE 2: NMR spectra of wild and mutant forms of CBM22-2. (A) Upfield NMR spectra of wild-type and mutant proteins. (B) NMR spectra of the wild-type protein and W53A in the region that contains the NH indole signals of tryptophan residues.

to −1.3 ppm, are derived from ring current-shifted protons which are in close contact with the π-orbitals of aromatic residues. The similarity of the spectra demonstrates that the hydrophobic core of the proteins is essentially intact and that the CBM22-2 mutants have all folded into a correct three-dimensional structure. This is particularly evident for mutants W53A, Y103A, and E138A, which have spectra very similar to that of the wild type. The spectra of R25A and Y136A are slightly different from those of the other proteins, suggesting that the amino changes in these proteins may have caused a slight reorientation of some aromatic residues in these mutants.

To investigate the role of the two highly conserved residues (Arg 25 and Glu 138) in CBM22-2, the three-dimensional structures of R25A and E138A were determined by X-ray crystallography. Details of the data and structure quality are given in Table 1. The data, presented in Figure 3, show that the R25A mutant causes a shift in the positions of two of the surface aromatic residues located in the cleft, Tyr 136 and Tyr 103. In contrast, the structure of E138A was essentially identical to that of the wild-type protein except for the removal of the carboxylate side chain.

Table 1: Refinement and Structure Quality Statistics for R25A and E138A

	R25A	E138A
data quality		
resolution (Å) (outer shell)	15−2.25 (2.33−2.25)	25−2.20 (2.28−2.20)
R_{merge}[a] (outer shell)	0.062 (0.367)	0.068 (0.35)
mean $I/\sigma I$ (outer shell)	29.7 (4.83)	24 (5.8)
completeness (%) (outer shell)	98.0 (97.0)	99.9 (100)
multiplicity (outer shell)	7.21 (6.35)	28.3 (25.5)
crystal parameters		
space group	$P4_32_12$	$P6_122$
cell dimensions	$a = b = 38.68$ Å, $c = 207.53$ Å	$a = b = 89.87$ Å, $c = 209.00$ Å
no. of moles per AU	1	2
refinement		
no. of protein atoms	1235	2447
no. of solvent waters	110	296
R_{cryst}	0.204	0.192
R_{free}	0.257	0.228
rms deviation for 1−2 bonds (Å)	0.009	0.011
rms deviation for 1−3 angles (deg)	1.259	1.377

[a] $R_{merge} = \sum_{hkl}\sum_i|I_{hkli} - \langle I_{hkl}\rangle|/\sum_{hkl}\sum_i\langle I_{hkl}\rangle$.

Affinity of CBM22-2 Mutants for Xylan. Xylan binding was initially assessed using AGE with both mutant and wild-type forms of the protein. The data (Figure 4) revealed that mutants Y103F, Y134A, and Y136F bind to oat spelt xylan with an affinity similar to that of native CBM22-2, while R25A, W53A, Y103A, Y136A, and E138A interact only weakly with the polysaccharide.

Quantitative determination of xylooligosaccharide affinity was achieved with ITC (Table 2). Y134A displayed an affinity for xylotetraose similar to that of native CBM22-2, while Y136A and E138A show no detectable binding. R25A, W53A, and Y103A have very low K_a values for the ligand, and Y103F and Y136F bound to the tetrasaccharide only 3- and 7-fold less tightly than the wild-type protein, respectively. The data clearly demonstrate that the cleft in CBM22-2 is the binding site for xylan. The large reduction in the level of xylotetraose binding when Arg 25, Trp 53, Tyr 103, Tyr 136, or Glu 138 was replaced with alanine implies that these amino acids interact directly with the ligand and/or are important for maintaining the structural integrity of the binding site. The thermodynamic parameters for ligand binding in Y134A were not significantly different from those of wild-type CBM22-2. Y103F displays a more favorable enthalpy than the wild type ($\Delta\Delta H = -1.7$ kcal mol^{-1}), but this is compensated by a significantly less favorable entropy ($T\Delta S = -2.3$ kcal mol^{-1}). In contrast, Y136F has a less

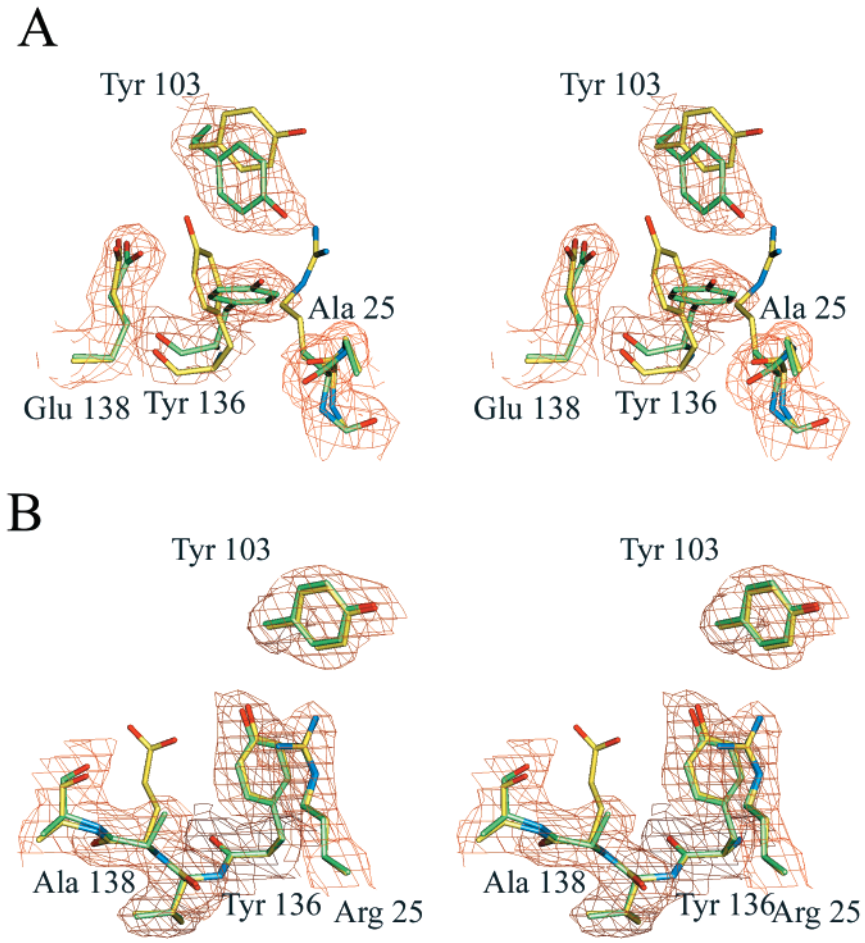

FIGURE 3: Comparison of the crystal structure of wild-type CBM22-2 and the R25A and E138A mutants. (A) Divergent (wall-eyed) stereo figure showing the R25A mutant. The electron density that is shown is a maximum-likelihood weighted $2F_{obs} - F_{calc}$ synthesis contoured at approximately 1σ (0.35 e/Å^3). Movement of the side chains of Tyr 103 and Tyr 136 accompanies the mutation. (B) Divergent (wall-eyed) stereo figure showing the E138A mutant. The electron density that is shown is a maximum-likelihood weighted $2F_{obs} - F_{calc}$ synthesis contoured at approximately 1σ (0.29 e/Å^3). Other than the loss of the side chain, the E138A mutant has no effects on the protein structure. The wild-type side chains are shown in yellow and the mutants in green.

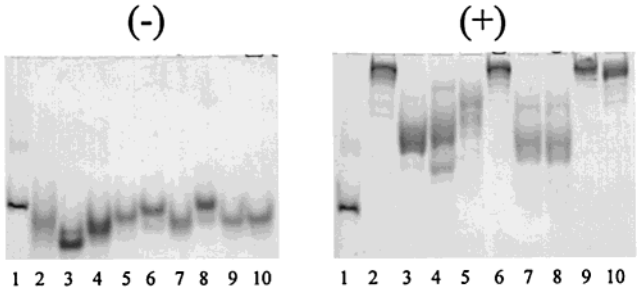

FIGURE 4: Affinity nondenaturing gel electrophoresis of wild-type and mutant forms of CBM22-2. Wild-type CBM22-2 (lane 2) and mutants R25A (lane 3), W53A (lane 4), Y103A (lane 5), Y134A (lane 6), Y136A (lane 7), E138A (lane 8), Y103F (lane 9), and Y136F (lane 10) were subjected to electrophoresis using 7.5% polyacrylamide gels in the absence (−) and presence (+) of 0.1% (w/v) oat spelt xylan. BSA was loaded in lane 1 as a negative control.

Table 2: Affinity of Wild-Type and Mutant Forms of CBM22-2 for Xylotetraose[a] Determined by ITC

protein	K_a ($\times 10^{-3}$ L mol^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})	n
wild-type	145 ± 2.4	−7.0 ± 0.01	−14.5 ± 0.5	−7.5 ± 0.5	1.0 ± 0.1
R25A	0.2	−3.1	—[b]	—[b]	—[b]
W53A	0.3	−3.5	—[b]	—[b]	—[b]
Y103A	0.5	−3.7	—[b]	—[b]	—[b]
Y134A	98.1 ± 4.0	−6.8 ± 0.02	−14.3 ± 0.3	−7.5 ± 0.3	1.0 ± 0.2
Y136A	ND[c]	—	—	—	—
E138A	ND[c]	—	—	—	—
Y103F	50.0 ± 4.7	−6.4 ± 0.06	−16.2 ± 0.1	−9.8 ± 0.2	1.1 ± 0.2
Y136F	21.1 ± 5.8	−5.9 ± 0.17	−13.3 ± 0.5	−7.4 ± 0.7	1.1 ± 0.1

[a] The stock concentration of xylotetraose used in the ITC titrations was 3 mM, except for the R25A, W53A, and Y103A mutants where 15 mM sugar was used. [b] Affinity too low to give accurate values for thermodynamic parameters (see the Results). [c] No binding detected.

favorable enthalpic contribution to binding than the native protein ($\Delta\Delta H = 1.2$ kcal mol^{-1}), but its entropy is almost unchanged. Accurate determination of enthalpy and entropy in the mutants which displayed low, but detectable, affinity by ITC (R25A, W53A, and Y103A) was not possible as the c values (product of the association constant K_a and the total macromolecule concentration in the cell) were all <1 (*39*).

Evidence That Trp 53 Interacts with Xylotetraose. To investigate whether specific signals in the NMR spectrum of the wild-type protein are influenced by the CBM binding to its ligand, CBM22-2 was titrated with xylotetraose. The NMR spectrum of the titrated protein showed that signals at −0.07, 7.3, and 10.3 ppm, which correspond to methyl, amide, and the indole NH of tryptophan, respectively, all moved (data not shown). The K_a calculated from the

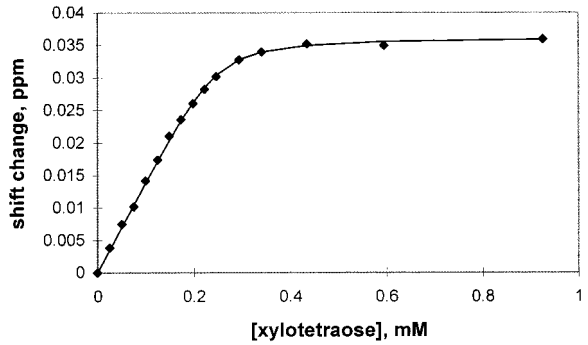

FIGURE 5: Effect of xylotetraose binding on the NMR spectrum of wild-type CBM22-2. The data points show experimental chemical shift changes of Trp HNϵ (10.3 ppm) on addition of xylotetraose. The curve shows the best fit binding isotherm, corresponding to a K_a of 1.3×10^5 M^{-1}.

movement of the indole NH signals was approximately 1.3×10^5 M^{-1} (Figure 5), which compares well with the value determined by ITC (Table 2). To investigate the identity of the tryptophan signal, the NMR spectra of the wild-type and mutant forms of CBM22-2 were recorded. The data, presented in Figure 2b, showed that the signal at 10.3 ppm was absent in the W53A mutant, indicating that the Trp 53 signal was altered when the protein interacted with xylotetraose. The change in chemical shift of Trp 53 HNϵ is only approximately 0.035 ppm, indicating that no change in hydrogen bonding to the side chain HNϵ occurs on binding.

Functional Importance of CBM22-1 in Xyn10B Catalytic Activity. Xyn10B contains two CBM22 modules which flank the GH10 xylanase catalytic module, termed CBM22-1 and CBM22-2, respectively. It has previously been shown that CBM22-2, when linked to the catalytic module, can potentiate enzyme activity against insoluble, but not soluble, xylan (*11*). To investigate the role of CBM22-1, the GH10 catalytic module, by itself or in its natural fusion with CBM22-1, was incubated with soluble oat spelt xylan, insoluble unsubstituted xylan, xylohexaose, xylopentaose, or xylotetraose and the released products were analyzed. The data, examples of which are presented in Figure 6, show that the rate of hydrolysis and the pattern of generated oligosaccharides were to some extent influenced by the presence or absence of CBM22-1. The only difference observed in the product profiles was a small increase in the amount of xylopentaose and xylohexaose with CBM22-1-GH10 compared to that with GH10 alone when the enzymes were incubated with insoluble unsubstituted xylan, although an explanation for this result is not readily apparent.

Functional Importance of CBM22-1 in Xyn10B Thermostability and Protease Resistance. Previous studies have indicated that removal of family 22 CBMs from their full-length enzymes can reduce the thermostability of the catalytic modules (*29*, *40*). To investigate further whether CBM22s influenced the stability of Xyn10B, the thermostability and proteinase sensitivity of CBM22-1-GH10, GH10, and GH10-CBM22−2 were compared. The data (Figure 7) showed that CBM22-1-GH10 was more thermostable and exhibited greater resistance to proteinase attack than the two other protein constructs. To investigate whether CBM22-1 could confer improved stability when not covalently attached, CBM22-1 and the GH10 catalytic module were mixed in equimolar concentrations and the properties of the enzyme

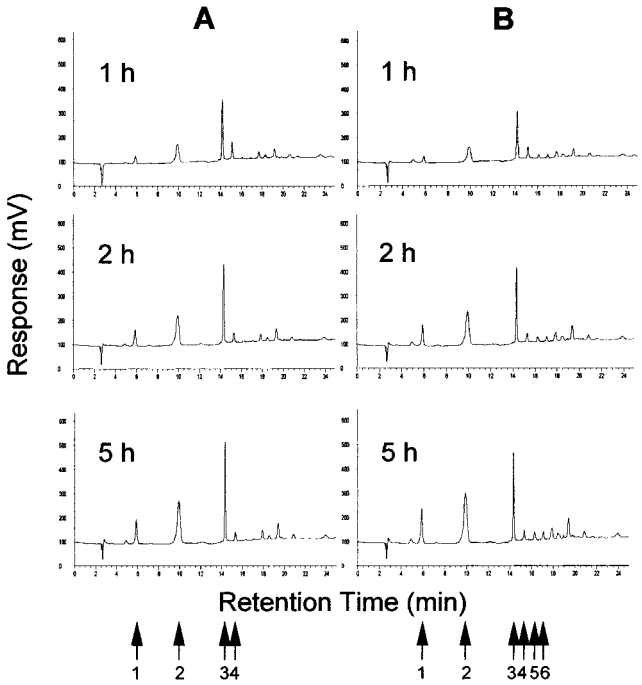

FIGURE 6: HPLC analysis of xylan hydrolysis by GH10CM and CBM22-1-GH10CM. In panels A and B, respectively, GH10CM and CBM22-1-GH10CM, matched in activity against PNPC, were incubated with 2% (w/v) insoluble unsubstituted xylan as described in Materials and Methods. Aliquots were removed at intervals, boiled for 10 min to inactivate the enzyme, and then subjected to HPLC. The positions at which xylose (1), xylobiose (2), xylotriose (3), xylotetraose (4), xylopentaose (5), and xylohexaose (6) were eluted from the HPLC column are indicated.

determined. Although nondenaturing polyacrylamide gel electrophoresis showed that the two proteins physically associated (data not shown), this did not lead to enhanced stability of the GH10 module.

DISCUSSION

Data presented in this study showed that replacing any of the five conserved residues located on the surface of the *C. thermocellum* Xyn10B CBM22-2 cleft generated mutant proteins that exhibited either no or greatly reduced affinity for xylan and xylotetraose. These results demonstrate that the ligand binding site is situated in the cleft. This conclusion is strongly supported by the NMR studies, which showed that the signal of the NH of the indole ring of Trp 53 titrated with xylotetraose, implying that the indole ring interacts with the tetrasaccharide. In contrast to the relatively deep cleft in CBM22-2, the ligand binding sites of the other two xylan-binding CBMs that have been characterized are quite different. The ligand binding sites of CBM2bs comprise a very shallow groove on the protein surface, while the lectin-like CBM13 consists of three homologous domains (α, γ, and δ) that contain individual saccharide binding sites positioned at 120° with respect to each other (*41*). A xylan chain occupies two of the three binding sites at any one time (*10*, *41*). The difference in the topology of the ligand binding sites of modules that interact with xylan is in sharp contrast with CBMs that bind to crystalline cellulose, which are very similar; they all contain a planar hydrophobic surface rich in aromatic amino acids that interact with the insoluble polysaccharide (*4*, *19*).

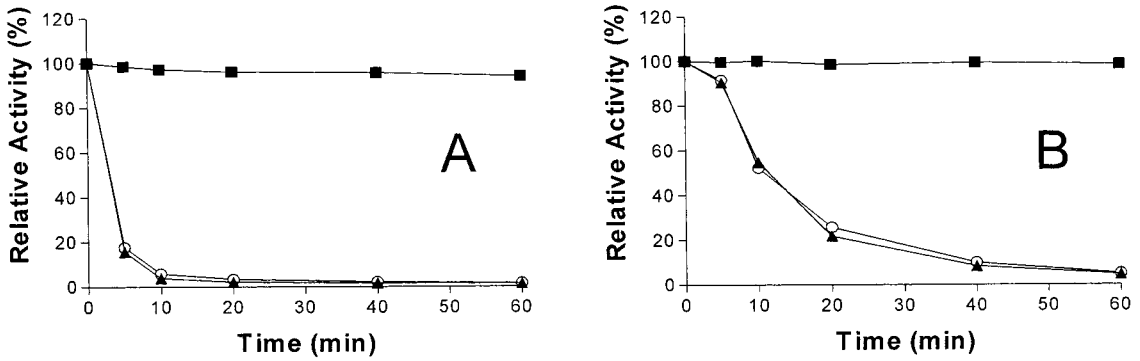

FIGURE 7: Thermostability and protease resistance of Xyn10B derivatives. GH10CM (▲), CBM22-1-GH10CM (■), and GH10CM-CBM22-2 (○) were incubated either at 65 °C (A) or with pancreatin at 37 °C (B) for up to 60 min and their remaining activities determined as described in Materials and Methods.

The site-directed mutagenesis studies showed that Tyr 134 does not interact with xylotetraose or xylan. This is consistent with the location of this residue at the very top of the cleft. In addition, this residue is not conserved among other CBM22 members. Changing any of the other three aromatic residues (Trp 53, Tyr 103, and Tyr 136) to alanine, however, caused a substantial reduction in the affinity for the ligand, consistent with the solvent-exposed position of Trp 53 and Tyr 103 in the cleft. These two residues are likely to form hydrophobic stacking interactions with adjacent sugar rings in ligand molecules. This view is supported by the small chemical shift changes of Trp 53 HNϵ on titration with xylotetraose, and by the observation that replacing Tyr 103 with phenylalanine did not significantly reduce ligand affinity, indicating that these aromatic residues do not form important hydrogen bonds with xylan or xylooligosaccharides.

The relative 150° aromatic ring orientation of Trp 53 and Tyr 103 is interesting in view of the three-fold helical structure of xylan, which requires the amino acids to be positioned at 120° with respect to one another for parallel stacking against the sugar rings. It is possible that the position of one or both of these amino acids changes when the ligand is bound so that parallel stacking interactions are made possible, or that when xylan binds to CBM22-2 it does not adopt its normal three-fold helical conformation. Alternatively, it may be that stacking between the aromatic residues and sugar rings is not perfectly parallel and that somewhat more angled interactions occur (*19*).

In contrast to Trp 53 and Tyr 103, the bulk of Tyr 136 is buried, with only its hydroxyl group exposed to solvent. This amino acid thus appears to interact with the ligand via hydrogen bonding rather than hydrophobic interactions. It is interesting to note, however, that changing Tyr 136 to alanine abolished ligand binding, but changing the residue to phenylalanine only reduced the affinity for xylotetraose 7-fold. These data suggest this residue has two roles in the function of CBM22-2; the phenolic ring plays a structural role in the binding site, while the hydroxyl group interacts with the ligand. This view is further supported by the thermodynamic data which show that the reduced affinity seen with Y136F is entirely due to a smaller enthalpic contribution to binding, consistent with the loss of a hydrogen bond. Replacing either Arg 25 or Glu 138 with alanine caused a substantial reduction in ligand affinity. It is possible that these data reflect the importance of hydrogen bonds between these amino acids and the carbohydrate, although the results are also consistent with a structural role for one or both of these residues.

To evaluate whether these hydrophilic amino acids, and the aromatic residues discussed above, play a structural role in CBM22-2, three-dimensional studies were conducted using both NMR spectroscopy and X-ray crystallography. The NMR spectroscopic data showed that W53A, Y103A, and E138A had three-dimensional structures very similar to that of the wild-type protein. In contrast, the NMR spectra of Y136A and R25A were somewhat different from native CBM22-2. X-ray crystallography of R25A and E138A confirmed the NMR data; removal of the glutamate side chain did not influence the structural integrity of the binding cleft, but substitution of the arginine side chain caused a structural reorientation of the two solvent-exposed aromatic residues. These structural studies strongly support the view that Trp 53, Tyr 103, and Glu 138 bind directly to the ligand, and each of these interactions is pivotal for efficient ligand binding. In contrast, Arg 25 has a structural role in the module, although its position on the surface of the cleft suggests that the large reduction in affinity caused by the R25A mutation is likely also to reflect the loss of hydrogen bonding between the ligand and the guanidino group.

The importance of hydrophobic interactions between the ligand and the two aromatic residues (Trp 53 and Tyr 103) is typical of many carbohydrate-active protein−carbohydrate interactions. In the CBMs that bind crystalline and amorphous cellulose, xylan, and starch, the stacking of aromatic amino acids with sugar rings plays a pivotal role in ligand binding. Removal of these aromatic residues, in each of the CBMs, greatly reduces, or abolishes, ligand affinity (*19*, *24*−*26*, *42*). The functional importance of polar amino acids that form hydrogen bonds with the ligand is less certain. In a recent study, Xie et al. (*26*) showed that the removal, from *Cellulomonas fimi* Xyn11A CBM2b-1, of three polar residues that hydrogen bond with ligand only resulted in a very modest reduction in affinity for xylan, but had a significant effect on the thermodynamic forces driving the interaction. The substantial reduction in enthalpy through the loss of hydrogen bonding was almost exactly compensated by an increase in entropy, which is due to an increase in the conformational mobility in the complex following removal of the strongly directional hydrogen bonds. Similarly, Kormos et al. (*25*) showed that removal of the majority of polar residues from *Ce. fimi* Cel9B CBM4-1, which also contains a relatively shallow cleft, that form hydrogen bonds with

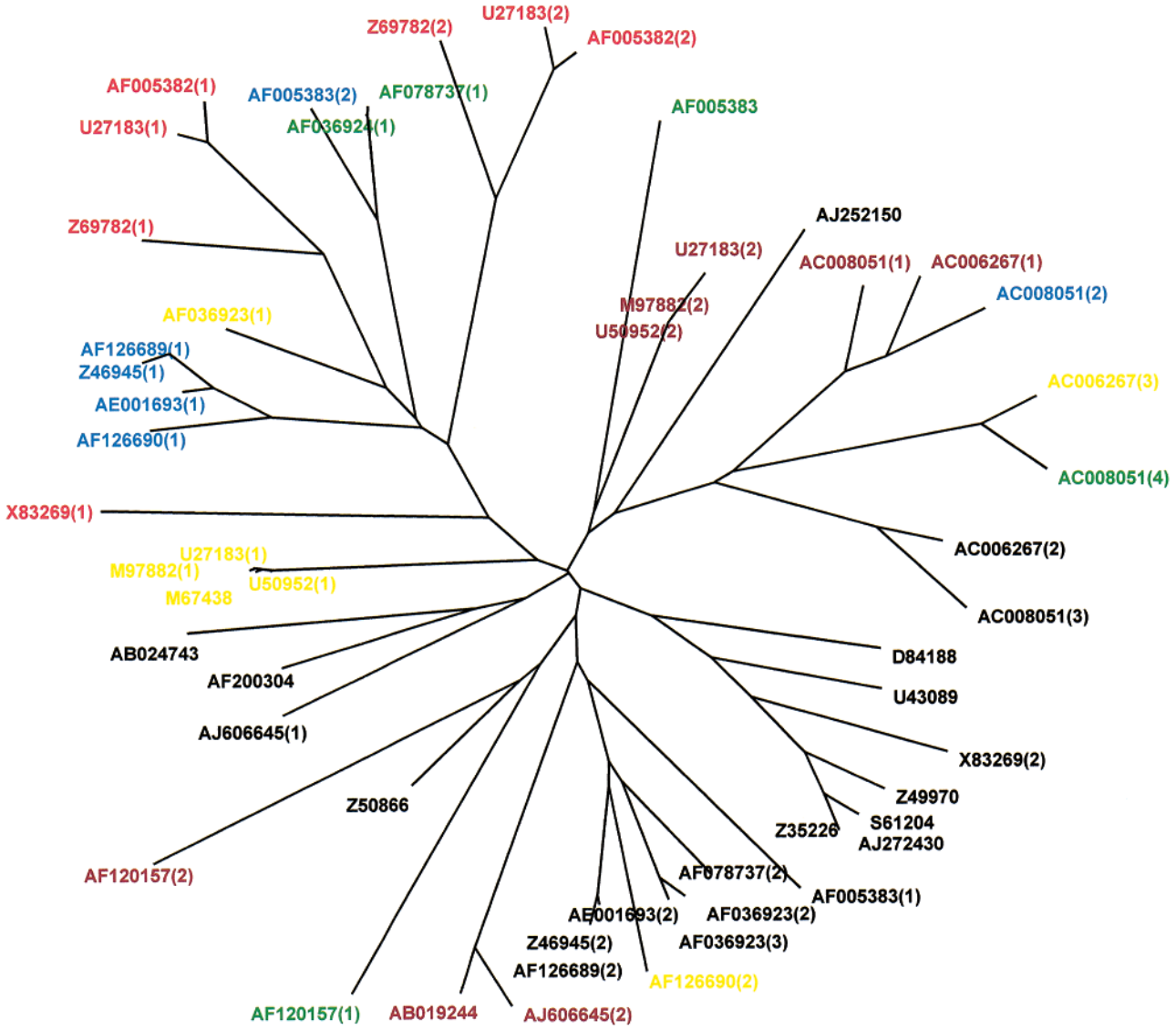

FIGURE 8: Unrooted phylogenetic tree of *C. thermocellum* Xyn10B CBM22-2 and other CBM22 proteins. The accession numbers of the proteins are given. Proteins labeled in black contain all five of the residues present in Ct Xyn10B CBM22-2 which are necessary for xylan binding. Proteins labeled in brown have only four of these residues, those in yellow only three, those in blue only two, those in green one, and those in red none. The numbers in parentheses indicate the order of the module in the full-length enzyme.

amorphous cellulose and β-glucan had little influence on ligand affinity. Removal of polar residues that are likely to interact with the target ligand from CBMs that bind to crystalline cellulose generally also has little influence on the overall affinity (*19*, *22*, *43*). In contrast to CBMs, hydrogen bonds play a critical role in the interaction of carbohydrates with proteins that contain more occluded binding clefts or pockets, exemplified by certain lectins (*44*). Thus, the importance of hydrogen bonds in protein−sugar interactions appears to lie in the structure of the ligand binding sites. In proteins with buried binding sites, the reduction in enthalpy through the loss of a protein−carbohydrate hydrogen bond cannot be compensated by increased conformational freedom of the ligand, due to steric hindrance imposed by the binding site. The view that the role of hydrogen bonds in the interaction of ligands with CBMs that have surface binding sites is less critical than in other protein−carbohydrate systems, is supported by the fact that in CBM22-2, which contains a relatively deep cleft, hydrogen bonds are more important in the capacity of the protein to bind its ligands than in CBMs where the polysaccharide binding site is more exposed (*19*, *25*, *26*). The reason for the smaller reduction in affinity in Y136F than in E138A could reflect the relative strength or number of the hydrogen bonds between the ligand and Tyr 136 and Glu 138. Currently, it is not possible to assess the importance of hydrogen bonding between Arg 25 and the ligand as this amino acid has been shown to play a pivotal role in maintaining the position of two critically important binding residues (Tyr 136 and Tyr 103). Although it is apparent that residues involved in hydrogen bonds with the sugar are less important in CBMs with a shallow binding cleft than in those with a deeper cleft, this is not the case with aromatic residues involved in hydrophobic stacking interactions, which have been shown to be pivotal in the binding of all CBMs studied to date. This is due to the fact that these hydrophobic interactions are primarily entropically driven, and thus, conformational restriction of the ligand imposed by the topology of the binding site is not important in the stacking of aromatic residues with the sugar rings.

Previous studies have shown that *C. thermocellum* Xyn10B and *Thermatoga maritma* Xyn10A both contain two copies of CBM22, only one of which was shown to bind xylan in each enzyme (*11*, *27*). Results presented in this study provide an explanation for the inability of some CBM22 modules to interact with the hemicellulose; they lack some or all of the amino acids that play a pivotal role in xylan binding. To predict how many of the other CBM22s are likely to bind xylan, the primary sequences of the 55 known CBM22 modules were compared. The data, presented in Figure 8, show that only 22 of the 55 CBM22s contain the five critical residues, Arg 25, Trp 53, Tyr 103, Tyr 136, and Glu 138, suggesting that numerous other CBM22 modules may not have the capacity to bind xylan. Phylogenetic analysis of the CBM22 sequences shows that the xylan-binding modules (proteins that contain the five residues shown to be important in xylan binding in *C. thermocellum* Xyn10B CBM22-2) are generally clustered in a single branch.

Although the function of CBM22s that contain all five xylan-binding residues is clear, the role of the other proteins in this family remains unclear. Data presented in this paper show that the presence of CBM22-1 increases the stability of the Xyn10B catalytic module; however, it is unlikely that this is the protein's primary role as its removal will inevitably disrupt the structure of the GH10 module, which directly abuts the family 22 CBM (*11*). It has been suggested that these modules (CBM22s that directly abut the catalytic module) may fulfill a function similar to that of CBM3c in family 9 cellulases, which feeds single glucan chains into the active site of the adjacent catalytic module, facilitating the processive action of this cellulase against crystalline substrates (*11*, *17*). Such a role for CBM22-1 of *C. thermocellum* Xyn10B is unlikely as removal of this module from the enzyme did not appear to significantly alter its activity or mode of action against either substituted or unsubstituted xylans or xylooligosaccharides. It is possible that CBM22s that lack the five xylan-binding residues bind to carbohydrates which have yet to be identified or that they bind at the interface between different polysaccharides, and such an interaction would not be evident when using isolated carbohydrates as the potential ligand. Indeed, a class of plant proteins known as expansins are believed to bind at the interface between cellulose microfibrils and hemicelluloses such as xyloglucan and glucomannan.

We have demonstrated that both hydrophobic interactions and hydrogen bonds play a key role in the binding of *C. thermocellum* Xyn10B CBM22-2 to xylan and xylotetraose. The fact that the five residues which are pivotal for the interaction of Xyn10B CBM22-2 with xylose polymers are not completely conserved in this family provides an explanation for the observation that not all proteins in this family bind xylan.

REFERENCES

1. Tomme, P., Warren, R. A. J., and Gilkes, N. R. (1995) *Adv. Microb. Physiol. 37*, 1−81.
2. Coutinho, P. M., and Henrissat, B. (2001) Carbohydrate-Binding Module Family server (http://afmb.cnrs-mrs.fr/~pedro/CAZY/cbm.html).
3. Coutinho, P. M., and Henrissat, B. (1999) in *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G. J., Henrissat, B., and Svensson, B., Eds.) pp 3−12, The Royal Society of Chemistry, Cambridge, U.K.
4. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) *EMBO J. 15*, 5739−5751.
5. Brun, E., Moriaud, F., Gans, P., Blackledge, M. J., Barras, F., and Marion, D. (1997) *Biochemistry 36*, 16074−16086.
6. Ponyi, T., Szabó, L., Nagy, T., Orosz, L., Simpson, P. J., Williamson, M. P., and Gilbert, H. J. (2000) *Biochemistry 39*, 985−991.
7. Simpson, P. J., Hefang, X., Bolam, D. N., Gilbert, H. J., and Williamson, M. P. (2000) *J. Biol. Chem. 275*, 41137−41142.
8. Abou Hachem, M., Nordberg Karlsson, E., Bartonek-Roxå, E., Raghothama, S., Simpson, P. J., Gilbert, H. J., Williamson, M. P., and Holst, O. (2000) *Biochem. J. 345*, 53−60.
9. Sun, J. L., Sakka, K., Karita, S., Kimura, T., and Ohmiya, K. (1998) *J. Ferment. Bioeng. 85*, 63−68.
10. Fujimoto, Z., Kuno, A., Kaneko, S., Yoshida, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2000) *J. Mol. Biol. 300*, 575−585.
11. Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M. A., Davies, G. J., and Fontes, C. M. G. A. (2000) *Biochemistry 39*, 5013−5021.
12. Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., and Williamson, M. P. (1998) *Biochem. J. 331*, 775−781.
13. Gill, J., Rixon, J. E., Bolam, D. N., McQueen-Mason, S. J., Simpson, P., Williamson, M. P., Hazlewood, G. P., and Gilbert, H. J. (1999) *Biochem. J. 342*, 473−480.
14. Din, N., Damude, H. G., Gilkes, N. R., Miller, R. C., Warren, R. A. J., and Kilburn, D. G. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11383−11387.
15. Southall, S. M., Simpson, P. J., Gilbert, H. J., and Williamson, M. P. (1999) *FEBS Lett. 447*, 58−60.
16. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) *Nat. Struct. Biol. 4*, 810−817.
17. Irwin, D., Shin, D.-H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998) *J. Bacteriol. 180*, 1709−1714.
18. Raghothama, S., Simpson, P. J., Szabó, L., Nagy, T., Gilbert, H. J., and Williamson, M. P. (2000) *Biochemistry 39*, 978−984.
19. McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A., and Kilburn, D. G. (2000) *Protein Eng. 13*, 801−809.
20. Johnson, P. E., Joshi, M. D., Tomme, P., Kilburn, D. G., and McIntosh, L. P. (1996) *Biochemistry 35*, 14381−14394.
21. Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J., and Williamson, M. P. (1999) *Structure 7*, 853−864.
22. Linder, M., Mattinen, M.-L., Kontteli, M., Lindeberg, G., Ståhlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G., and Annila, A. (1995) *Protein Sci. 4*, 1056−1064.
23. Din, N., Forsythe, I. J., Burntnick, L. D., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. J., and Kilburn, D. G. (1994) *Mol. Microbiol. 11*, 747−755.
24. Nagy, T., Simpson, P. J., Williamson, M. P., Hazlewood, G. P., Gilbert, H. J., and Orosz, L. (1998) *FEBS Lett. 429*, 312−316.
25. Kormos, J., Johnson, P. E., Brun, E., Tomme, P., McIntosh, L. P., Haynes, C. A., and Kilburn, D. G. (2000) *Biochemistry 39*, 8844−8852.
26. Xie, H., Bolam, D. N., Nagy, T., Szabó, L., Cooper, A., Simpson, P. J., Lakey, J. H., Williamson, M. P., and Gilbert, H. J. (2001) *Biochemistry 40*, 5700−5707.
27. Meissner, K., Wassenberg, D., and Liebl, W. (2000) *Mol. Microbiol. 36*, 898−912.
28. Sunna, A., Gibbs, M. D., and Bergquist, P. L. (2000) *Biochem. J. 346*, 583−586.
29. Fontes, C. M. G. A., Hazlewood, G. P., Morag, E., Hall, J., Hirst, B. H., and Gilbert, H. J. (1995) *Biochem. J. 307*, 151−158.
30. Millar, G. L. (1959) *Anal. Chem. 31*, 426−428.
31. Pace, C. N., Vadjos, F., Fee, L., Grimsley, G., and Gray, T. (1995) *Protein Sci. 4*, 2411−2423.

32. Ghangas, G. S., Hu, Y.-J., and Wilson, D. B. (1989) *J. Bacteriol. 171*, 2963−2969.
33. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
34. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
35. Navaza, J. (1994) *Acta Crystallogr. 50*, 157−163.
36. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D53*, 240−255.
37. Thompson, J. D., Higgins, D. G., and Gibson, T. L. (1994) *Nucleic Acids Res. 22*, 4673−4680.
38. Strimmer, K., and von Haeseler, A. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 6815−6819.
39. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L.-N. (1989) *Anal. Biochem. 179*, 131−137.
40. Winterhalter, C., Heinrich, P., Candussio, A., Wich, G., and Liebl, W. (1995) *Mol. Microbiol. 15*, 431−444.
41. Boraston, A. B., Tomme, P., Amandoron, E. A., and Kilburn, D. G. (2000) *Biochem. J. 350*, 933−941.
42. Williamson, M. P., Le Gal-Coëffet, M.-F., Sorimachi, K., Furniss, C. S. M., Archer, D. B., and Williamson, G. (1997) *Biochemistry 36*, 7535−7539.
43. Simpson, H. D., and Barras, F. (1999) *J. Bacteriol. 181*, 4611−4616.
44. Lemieux, R. U. (1989) *Chem. Soc. Rev. 18*, 347−374.
45. Esnouf, R. M. (1997) *J. Mol. Graphics 15*, 133−138.
46. Merritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D50*, 869−873.

BI0106742