doi:10.1016/j.jmb.2003.12.081 J. Mol. Biol. (2004) 337, 417–426

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Ligand-mediated Dimerization of a Carbohydrate-binding Module Reveals a Novel Mechanism for Protein–Carbohydrate Recognition

James Flint[1], Didier Nurizzo[2], Stephen E. Harding[3], Emma Longman[3] Gideon J. Davies[2], Harry J. Gilbert[1] and David N. Bolam[1]*

[1]*School of Cell and Molecular Biosciences, University of Newcastle upon Tyne, The Agriculture Building Newcastle upon Tyne NE1 7RU, UK*

[2]*Structural Biology Laboratory Department of Chemistry University of York, Heslington York YO10 5YW, UK*

[3]*National Centre for Macromolecular Hydrodynamics, School of Biosciences, University of Nottingham, Sutton Bonington Loughborough LE12 5RD, UK*

**Corresponding author*

The structural and thermodynamic basis for carbohydrate–protein recognition is of considerable importance. NCP-1, which is a component of the *Piromyces equi* cellulase/hemicellulase complex, presents a provocative model for analyzing how structural and mutational changes can influence the ligand specificity of carbohydrate-binding proteins. NCP-1 contains two "family 29" carbohydrate-binding modules designated CBM29-1 and CBM29-2, respectively, that display unusually broad specificity; the proteins interact weakly with xylan, exhibit moderate affinity for cellulose and mannan, and bind tightly to the β-1,4-linked glucose-mannose heteropolymer glucomannan. The crystal structure of CBM29-2 in complex with cellohexaose and mannohexaose identified key residues involved in ligand recognition. By exploiting this structural information and the broad specificity of CBM29-2, we have used this protein as a template to explore the evolutionary mechanisms that can lead to significant changes in ligand specificity. Here, we report the properties of the E78R mutant of CBM29-2, which displays ligand specificity that is different from that of wild-type CBM29-2; the protein retains significant affinity for cellulose but does not bind to mannan or glucomannan. Significantly, E78R exhibits a stoichiometry of 0.5 when binding to cellohexaose, and both calorimetry and ultracentrifugation show that the mutant protein displays ligand-mediated dimerization in solution. The three-dimensional structure of E78R in complex with cellohexaose reveals the intriguing molecular basis for this "dimeric" binding mode that involves the lamination of the oligosaccharide between two CBM molecules. The 2-fold screw axis of the ligand is mirrored in the orientation of the two protein domains with adjacent sugar rings stacking against the equivalent aromatic residues in the binding site of each protein molecule of the molecular sandwich. The sandwiching of an oligosaccharide chain between two protein modules, leading to ligand-induced formation of the binding site, represents a completely novel mechanism for protein–carbohydrate recognition that may mimic that displayed by naturally dimeric protein–carbohydrate interactions.

Keywords: carbohydrate binding; dimerization; cellulose recognition; carbohydrate-binding modules; X-ray crystallography

Introduction

Protein–carbohydrate recognition plays a pivotal role in several key biological processes including host–pathogen recognition, cell–cell communication, defence, protein trafficking and the recycling of carbon though the degradation of the plant cell-wall. Plant cell-wall hydrolases, derived

Abbreviations used: HEC, hydroxyethylcellulose; CGM, carob galactomannan; KGMh, high-viscosity konjac glucomannan; CBM, carbohydrate-binding module; CBM29, proteins belonging to family 29 of the carbohydrate-binding modules; NCP, non-catalytic protein; ITC, isothermal titration calorimetry.

E-mail address of the corresponding author: d.n.bolam@ncl.ac.uk

from aerobic organisms, have a modular structure in which non-catalytic carbohydrate-binding modules (CBMs) target the enzymes to specific polysaccharides.[1] The resultant increase in the concentration of the hydrolases greatly enhances the catalytic efficiency of these enzymes on insoluble and recalcitrant substrates.[1,2] CBMs have been grouped into a number of discrete families based upon amino acid sequence similarity†.[3] Within several of these sequence-based families, subtle differences in structure leads to diverse ligand specificity.[4–7] Thus, CBMs in complex with their target sugars present excellent systems for dissecting the molecular determinants that define the structural basis for protein–carbohydrate recognition.

Three-dimensional (3D) structures of CBMs have shown that the topography of the binding sites in these proteins reflects the macroscopic nature of the target ligand. This has led to the classification of these modules into types A, B or C.[1] Type A CBMs, which interact with crystalline cellulose, display a planar hydrophobic surface that interacts with adjacent chains of the 010 face of the crystal lattice.[8,9] Type B CBMs bind to less-structured plant structural polysaccharides, accommodating single sugar chains in a groove that extends along the complete length of the protein and are the commonest class of CBMs.[5,6,10–12] Type C CBMs contain small binding sites that interact with mono- or disaccharides.[13,14] A characteristic of CBMs that distinguishes them from lectins is a scarcity of hydrogen bonds between these proteins and their target ligands; instead, binding is dominated by hydrophobic interactions.[4,15–21] The lack of hydrogen bonds facilitates the accommodation of heterogeneous ligands in which the backbone saccharide polymer is often decorated with an array of different sugars and organic acids,[22] and may contribute to the necessarily dynamic nature of the interaction.

CBM family 29 contains two proteins that represent particularly interesting models for studying protein–carbohydrate recognition. The two CBM29 modules are present in a non-catalytic protein (NCP-1)[23] that is a component of the multienzyme cellulase/hemicellulase complex of the anaerobic fungus *Piromyces equi*, one of the most active plant cell-wall-degrading organisms known.[24] The N and C-terminal copies of CBM29 in NCP-1 are defined as CBM29-1 and CBM29-2, respectively. Both modules interact with the same array of polysaccharides; they are highly discriminatory for β-1,4-linked gluco and manno-configured ligands, displaying highest affinity for glucomannan.[23] Recently, the crystal structure of CBM29-2 in complex with mannohexaose and cellohexaose has been solved.[12] The data reveal a beta-jelly-roll topology, with an extended binding groove on the concave surface that can accommodate six sugar residues, typical of type B CBMs, whilst recognition of both manno and gluco-configured oligo- and polysaccharides is conferred by the plasticity of the direct interactions between the protein and the axial and equatorial 2-hydroxyl groups of mannose and glucose, respectively. The O6 of alternate sugar molecules are solvent-exposed, explaining how the protein is able to accommodate α-1,6-galactosyl side-chains.[12]

The broad specificity of CBM29-2 presents an excellent model for dissecting the molecular basis for the evolution of novel specificities within carbohydrate-binding proteins. Here, we show that replacing Glu78, which makes an important hydrogen bond with both gluco and manno-configured ligands, with an arginine residue has a profound effect on the ligand specificity of the CBM. The E78R mutant recognizes cellulose and cellohexaose but does not bind to polymers of mannose, and displays ligand-mediated dimerization in solution. The 3D structure of the mutant protein in complex with cellohexaose reveals the intriguing molecular basis for this "dimeric" binding mode involving the lamination of cellohexaose between two molecules of E78R. The sandwiching of an oligosaccharide chain between two protein molecules represents a novel molecular lamination mechanism for protein–carbohydrate recognition, which may mimic that displayed by natural dimeric carbohydrate-binding modules.

Results and Discussion

Basis for mutant construction

The crystal structure of CBM29-2 in complex with cellohexaose and mannohexaose[12] showed that the protein contains six sugar-binding subsites, and the $O^{\varepsilon 1}$ and $O^{\varepsilon 2}$ of Glu78 make a hydrogen bond to the O3 of either glucose or mannose bound at subsite 4. To determine the importance of these hydrogen bonds, Glu78 was replaced with alanine (E78A), and to explore the effect of introducing a radical amino acid change at this location, Glu78 was substituted with arginine to create E78R. The two mutant proteins were expressed in *Escherichia coli* and purified to electrophoretic homogeneity.

Ligand specificity of CBM29-2 E78A

The capacity of the mutant proteins to bind to polysaccharides was evaluated by affinity gel electrophoresis and compared to the respective wild-type protein (Table 1). The E78A mutant displayed a tenfold reduction in affinity for gluco and/or manno-configured polysaccharides, compared to wild-type CBM29-2, although the mutation did not influence the relative specificity of the protein for its ligands. These data indicate that the hydrogen bonds between the side-chain of Glu78 and the O3 of either a mannose or glucose moiety positioned

† http://afmb.cnrs-mrs.fr/CAZY/

Table 1. Ligand specificity of CBM29 wild-type and mutants determined by affinity gel electrophoresis

	K_A (% c^{-1})[a]		
Ligand	CBM29-2	E78R	E78A
HEC	15.4	4.21	1.39
CGM	45.9	0.28	4.91
KGMh	141	5.07	8.37

[a] The term % c^{-1} defines the association constant (K_A) as the reciprocal of the polysaccharide concentration expressed as a percentage.

in subsite 4 make a significant contribution to the overall affinity of the protein for its target ligands equating to a loss in binding energy of 1.4 kcal mol^{-1}, consistent with the strength of a direct hydrogen bond (1 cal = 4.184 J).[25] Analysis of E78A-ligand interactions by isothermal titration calorimetry (ITC) was not possible, as binding was too weak to quantify accurately (data not shown). These data are consistent with previous studies that have utilised the crystal structure of CBM–ligand complexes to examine the importance of direct hydrogen bonds in defining the overall affinity of these macromolecular associations. In contrast to type A CBMs, where hydrogen bonds appear to play a minor role in ligand binding,[18,21,26] substitution of potential direct hydrogen-bonding residues in type B CBMs (such as CBM4, CBM15 and CBM17), typically results in a sixfold to 35-fold reduction in binding.[5,27,28]

CBM29-2 E78R binds ligands as a dimer

Analysis of the ligand specificity of the E78R mutant by affinity gel electrophoresis (Table 1) shows that the protein displays no significant affinity for carob galactomannan, interacts with konjac glucomannan only weakly, but has substantial affinity for hydroxyethylcellulose (HEC) compared to wild-type CBM29-2. ITC was used to determine the stoichiometry of binding and the thermodynamic forces that drive the interaction of E78R with its ligands. As the affinity of E78R for oligosaccharides was relatively low, ITC was carried out at a temperature lower than the standard 25 °C to increase the c values. The data at both 25 °C and 5 °C (Figure 1 and Table 2) confirmed that the protein has significant affinity for substituted cellulose and cellohexaose, but does not bind to substituted mannans or mannohexaose. Most intriguingly, the stoichiometry of binding of E78R to cellohexaose or HEC is markedly different from that of wild-type CBM29-2. At saturation, native CBM29-2 and the E78R mutant bind to HEC at a ratio of protein to glucose molecules of 1:18 and 1:9, respectively. The wild-type CBM binds to cellohexaose in the expected ratio of 1:1, whilst the mutant displays a stoichiometry of 0.5 for the oligosaccharide. These data indicate that HEC and cellohexaose interact with a dimeric form of E78R.

To further probe the oligomeric state of E78R, the mutant protein was subjected to sedimentation equilibrium ultracentrifugation in the presence and in the absence of cellohexaose. From analysis of the equilibrium concentration data (Figure 2) using MSTAR, the molecular mass (M) in the absence of ligand was 17,400(±1000) Da, which is in excellent agreement with the protein sequence derived-molecular mass of E78R in its monomeric form (17,269 Da). In the presence of cellohexaose, however, M increased to 28,700(±1500) Da, approaching the dimeric value of 34,538 Da. NONLIN analysis (Figure 2) of the same equilibrium concentration data yielded a mass of 16,700(±1200) and 30,000(±1600) in the absence and in the presence of the ligand, respectively, and a K_D of ~120 μM for dimerization corresponding to an association constant K_A of $\sim 8 \times 10^3$ M^{-1}, a value comparable with the affinity of the protein for cellohexaose as determined by ITC (Table 2).

In general, monomeric carbohydrate-binding proteins interact with their oligosaccharide ligands with a stoichiometry of 1:1. Similarly, the subunits of oligomeric forms of these proteins also have a single ligand-binding site, although the interaction of these discrete binding sites with a common complex carbohydrate ligand results in a substantial increase in affinity (or avidity).[29] Deviation from this 1:1 stoichiometry, however, has been reported.

Table 2. Ligand specificity of CBM29-2 wild-type and mutants determined by isothermal titration calorimetry

Protein	Ligand[a]	Temperature (°C)	K_A (×10^3 M^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})	n
Wild-type	Mannohexaose	25	2.5 (±0.01)	−4.6 (±0.01)	−8.7 (±0.10)	−4.0 (±0.01)	1.48 (±0.14)
Wild-type	CGM	25	92.1 (±10.81)	−6.8 (±0.07)	−13.5 (±1.32)	−6.8 (±1.31)	1.02 (±0.07)
Wild-type	HEC	25	12.1 (±0.21)	−5.7 (±0.02)	−10.9 (±0.73)	−5.3 (±0.72)	0.95 (±0.03)
Wild-type	Cellohexaose	25	15.9 (±0.24)	−5.8 (±0.01)	−8.1 (±0.44)	−2.2 (±0.41)	1.17 (±0.02)
Wild-type	Cellohexaose	5	49.6 (±2.26)	−6.0 (±0.05)	−7.4 (±0.20)	−1.5 (±0.1)	1.02 (±0.02)
E78R	Mannohexaose	25	ND[b]	–	–	–	–
E78R	CGM	25	ND[b]	–	–	–	–
E78R	HEC	25	3.9 (±0.31)	−4.9 (±0.05)	−27.1 (±1.13)	−22.2 (±1.22)	0.37 (±0.02)
E78R	Cellohexaose	25	1.9 (±0.31)	−4.5 (±0.03)	−20.5 (±0.64)	−16.0 (±0.71)	0.53 (±0.02)
E78R	Cellohexaose	5	9.2 (±0.50)	−5.0 (±0.03)	−17.1 (±1.79)	−12.1 (±1.81)	0.53 (±0.01)

[a] The ligands CGM and HEC are carob galactomannan and hydroxyethylcellulose, respectively.
[b] ND, no binding detectable.

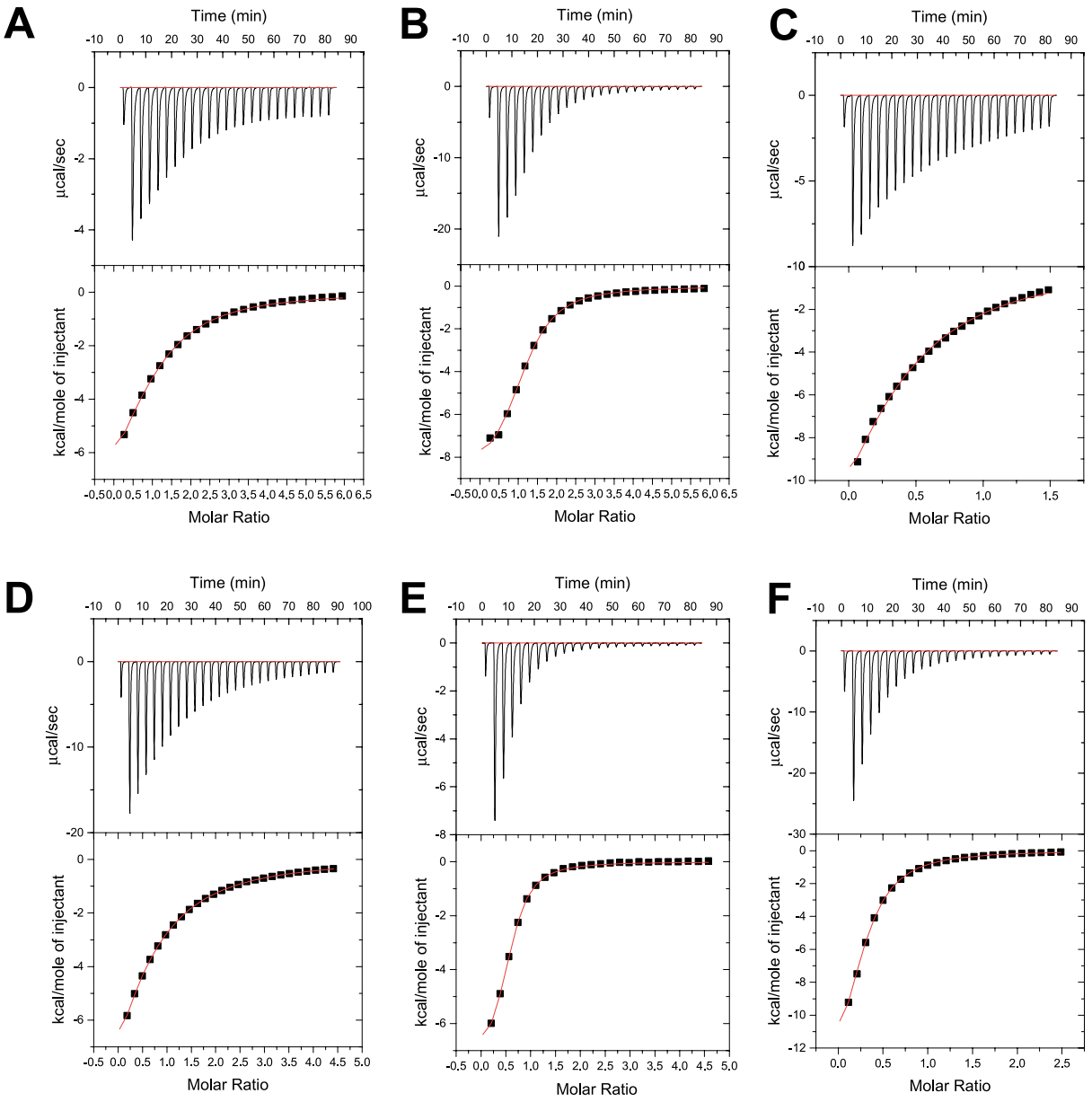

Figure 1. Isothermal titration calorimetry of wild-type CBM29-2 and E78R with HEC and cellohexaose. A and C, Titrations of CBM29-2 (100 μM) and E78R (400 μM) with hydroxyethylcellulose, respectively; B, D, E and F, titrations of these proteins with cellohexaose. The proteins and their concentrations were as follows: B, CBM29-2 (300 μM); D, E78R (400 μM); E, CBM29-2 (200 μM); F, E78R (760 μM). In A–D the titrations were carried out at 25 °C, while in E and F the titrations were carried out at 5 °C.

Thus, lectins such as ricin toxin B, GNA and tachylectin have multiple ligand-binding sites per subunit.[29] Examples where the ligand-binding site is at the interface between two distinct subunits in carbohydrate-binding proteins are rare, although this does occur in the wheatgerm agglutinin lectin WGA.[30] The protein comprises two subunits and contains four ligand-binding sites that are present at the interface between the subunits and thus amino acids from the different subunits contribute to the individual sugar-binding sites. The oligomeric structure of WGA is independent of the presence of its ligand, and thus the "pre-formed" carbohydrate-binding sites are a consequence of dimerization, not a result of ligand binding. In contrast, the ultracentrifugation data of E78R point to a ligand-mediated dimerization of the protein, implying that the oligosaccharide is acting as a molecular bridge between two molecules of CBM29-2, leading to a tri-molecular complex.

Crystal structure of E78R in complex with cellohexaose

Initial attempts to crystallize the unliganded form of E78R were unsuccessful. Crystals of the protein were, however, obtained in the presence of cellohexaose and data were collected to 1.5 Å at

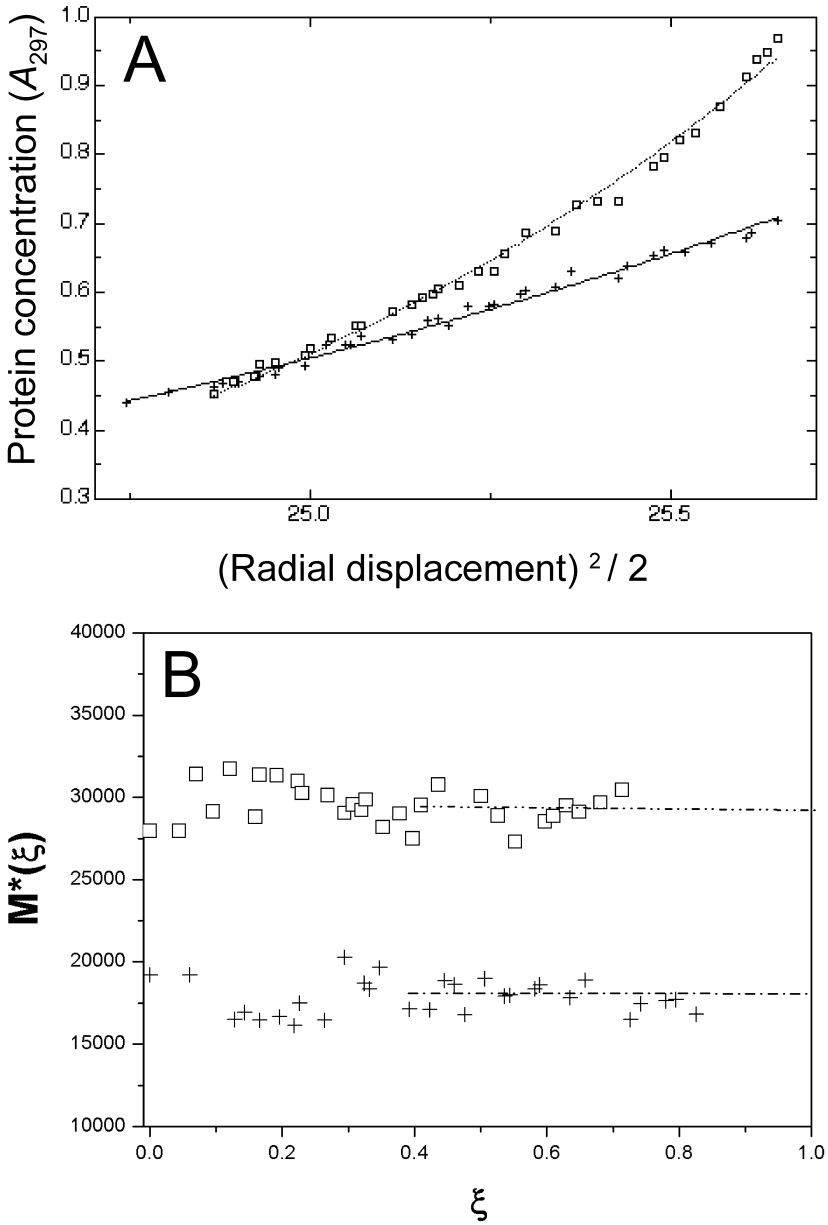

Figure 2. Analytical ultracentrifugation of E78R in the presence and in the absence of cellohexaose. A, The NONLIN fit of the equilibrium data plotted as protein concentration in absorbance units against (radial displacement squared)/2. From the fit of the data, the weight average molecular masses were obtained for E78R with (□) and without (+) 1 mM cellohexaose. B, the MSTAR extrapolation to estimate the weight average molecular mass ($M*(\zeta \rightarrow 1)$) for E78R with (□) and without (+) 1 mM cellohexaose. ξ is a normalized radial displacement square parameter $(r^2 - a^2)/(b^2 - a^2)$, at a given position in the ultracentrifuge cell, where r is the radial displacement at a given position in the ultracentrifuge cell with a and b the corresponding radial displacements at the cell meniscus and base, respectively. The rotor speed was 17,000 rpm and the temperature was 25.0 ° C (see Materials and Methods).

the ESRF (Grenoble France, see Materials and Methods). The 3D structure of the mutant protein in complex with cellohexaose (Figure 3(a)) is entirely consistent with the solution calorimetry and ultracentrifugation results and reveals the intriguing molecular basis for the "dimeric" binding mode, which involves the "lamination" of cellohexaose between two molecules of E78R (Figure 3(b)). The 2-fold screw axis of the ligand, cellohexaose, is mirrored in the orientation of the two protein domains. Adjacent sugar rings thus stack against equivalent aromatic residues in the binding site of each protein molecule of the dimeric sandwich. Remarkably, the ligand is rotated 180° relative to the conformation of cellohexaose bound to wild-type CBM29-2, with the aromatic residues packing against the "β-face" of the sugar rings, while the same amino acid residues in the native CBM29-2 module interact with the "α-face" (defined by the orientation of O1) of the glucopyranose units (Figure 3).[12] It is this interaction with the β-face of the ligand that appears to confer the newly engineered steric discrimination of E78R against the axial O2 of mannosides, since a protruding axial hydroxyl group would make steric clashes with the aromatic rings present in the binding site.

Cellohexaose is located in a more "shallow" location in the mutant protein compared to wild-type CBM29-2. A contribution to this change in location of the oligosaccharide is the bulky side-chain of arginine compared to glutamate, which would most likely clash with the ligand if it occupied the same position in the mutant and wild-type proteins. The displacement of the ligand out of the binding cleft and the exposure of the β-face of the sugar rings to the aromatic residues would result in relatively weak hydrophobic interactions with a monomer of E78R due to the increased distance between the interacting residues and the

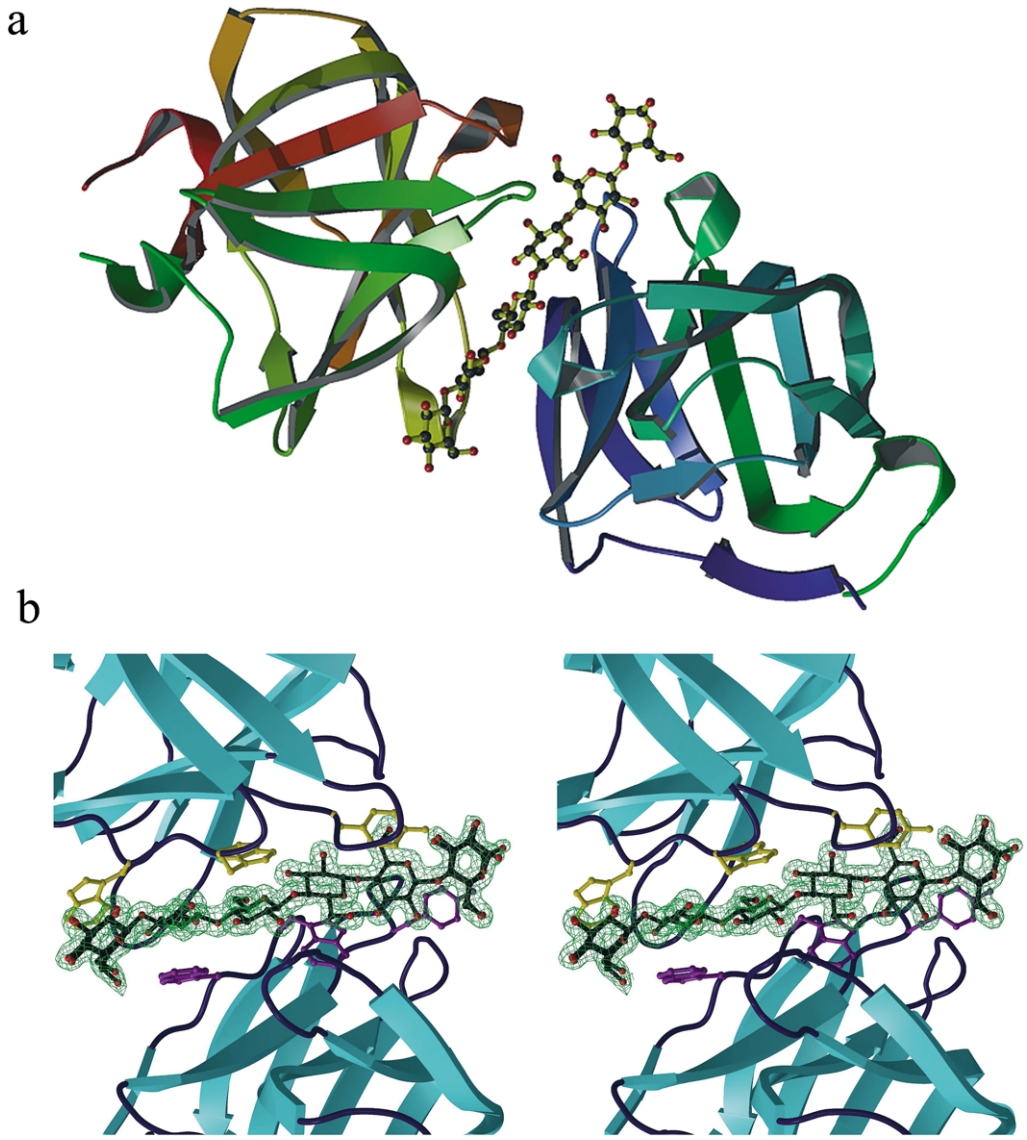

Figure 3. Crystal structure of E78R in complex with cellohexaose. (a) The crystal structure of the CBM29-2 mutant E78R in complex with cellohexaose. The molecules are color-ramped from N to C terminus, whilst the ligand is shown in ball-and-stick representation. (b) The observed electron density for the cellohexaose ligand together with the complementary aromatic interface provided by two tryptophan residues and a tyrosine residue from each protein monomer (purple and yellow licorice). This Figure was drawn with BOBSCRIPT.[41]

relatively hydrophilic nature of this sugar face. The position of the oligosaccharide higher up the binding cleft, however, enables the aromatic residues of a second molecule of E78R to interact with alternate sugar residues, and thus the dimeric conformation of the CBM29 mutant compensates for relatively weak interactions between single aromatic amino acids and sugar rings by increasing the extent of these hydrophobic interactions. Although the displacement of the ligand in the binding site of E78R places several of the polar groups that interact with cellohexaose in the wild-type protein, out of hydrogen bonding range of the ligand in the E78R complex, the glucose polymer does form alternative hydrogen bonds with the mutant protein (Figure 4). Thus the $N^{\varepsilon 2}$ of Gln116a (the two molecules of E78R in the complex designated a and b, respectively) forms a hydrogen bond with the O2 of Glc2 (Glc1 is the sugar at the reducing end of cellohexaose), while Gln116b makes a hydrogen bond with the O2 of Glc3. In wild-type CBM29-2 the N^{ε} and O^{ε} of Gln116 interact with the endocyclic oxygen atom and O6 of Glc3, respectively. The guanidino group of Arg78a and Arg78b form hydrogen bonds with the O6 of Glc3 and Glc4, respectively (in the wild-type protein both $O^{\varepsilon 1}$ and $O^{\varepsilon 2}$ of Glu78 interact with the O3 of Glc4), while the side-chain of Arg112a and Arg112b make hydrogen bonds with the O2 of Glc4 and Glc5, respectively. The carbonyl group of Gly25a and Gly25b form hydrogen bonds with the O6 of Glc4 and Glc5, respectively, while in the wild-type protein this amino acid does not interact with ligand. Thus, the total number of hydrogen bonds between the two E78R proteins and cellohexaose is nine, while wild-type CBM29-2 makes

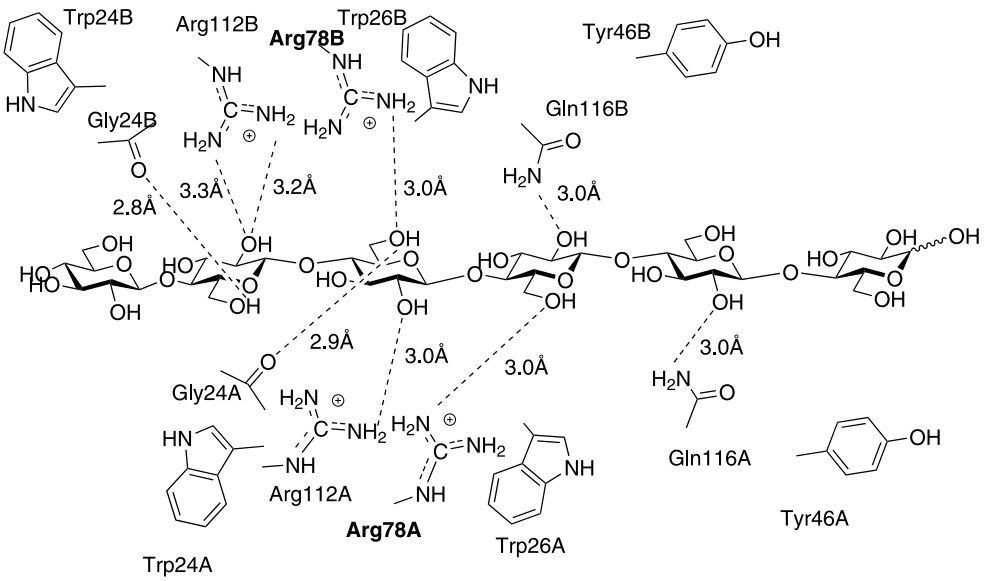

Figure 4. A diagram of the interactions of CBM29-2 E78R with cellohexaose.

ten hydrogen bonds with the hexasaccharide. In addition to the carbohydrate–CBM interactions, there are at least one water-mediated and three direct hydrogen bonds between the two protein molecules. The small number of direct interactions between the two E78R molecules explains why this protein does not form a dimer in the absence of ligand. The association of two CBM molecules with a common ligand, however, stabilizes the weak hydrogen bonds between the two proteins, and thus this co-operative network of hydrogen bonds and hydrophobic stacking interactions maintains the tri-molecular structure of the complex.

Although E78R is an artificially created mutant and thus not a natural molecule, the ligand-mediated assembly of the carbohydrate-binding site at the interface between the homodimer of E78R represents a novel mechanism of carbohydrate–protein recognition. One of the interesting questions raised by the interaction of E78R with its ligand is whether similar mechanisms for protein–carbohydrate recognition exist in other systems. Recently, a protein–carbohydrate interaction that bears some similarity to the E78R dimer–cellohexaose complex has been reported. CBM27 from *Thermotoga maritima* Man5A binds tightly to mannohexaose with a stoichiometry of 1:1. At a high concentration of protein, a second CBM molecule binds very weakly to the CBM27–ligand complex to form a protein homodimer bound to mannohexaose,[31] and this tri-molecular complex may resemble that of E78R bound to cellohexaose.

Conclusions

We believe that the cellohexaose-mediated assembly of the carbohydrate recognition site in the CBM29-2 mutant E78R represents a novel mechanism by which proteins recognize carbohydrates through molecular lamination. The location of the oligosaccharide at the interface between the two protein modules imposes additional steric constraints that change the ligand specificity of CBM29-2.

Materials and Methods

Expression and purification of CBM29-2 and E78R

Construction of the pET22b derivative pVM2 encoding CBM29-2 has been described.[23] CBM29-2 comprises residues 335–478 of *P. equi* NCP-1 and contains a C-terminal His_6 tag.[23] Standard culture conditions were used to express wild-type CBM29-2 and its mutants E78R and E78A in *E. coli* BL21(DE3):pLysS (Novagen). The proteins were purified to electrophoretic homogeneity by metal ion affinity chromatography.[12] For crystallographic purposes, E78R was further purified by anion-exchange and gel-filtration chromatography, essentially as described.[12]

Construction of CBM29 mutants

QuikChange site-directed mutagenesis (Stratagene) was carried out according to the manufacturer's instructions. The plasmid pVM2 was used as the template DNA to obtain mutants of CBM29-2 and the primers employed in the mutagenesis PCRs were as follows:

E78R

5′-GGTTAAGATTCTTGTTCGAAACTCTGAAGC-3′

5′-GCTTCAGAGTTTCGAACAAGAATCTTAACC-3′

E78A

5′-GGTTAAGATTCTTGTTGCGAACTCTGAAGC-3′

5′-GCTTCAGAGTTCGCAACAAGAATCTTAACC-3′.

Codons and anticodons for the introduced amino acids are underlined. The complete sequence of the DNA encoding the mutants was determined to confirm that only the desired mutation had been introduced.

Determination of ligand binding

Ligand binding was determined by both affinity gel electrophoresis and ITC. Affinity gel electrophoresis was performed as described[23] using HEC, high-viscosity konjac glucomannan and carob galactomannan (Megazyme International) as the ligands, which were included in the gels at concentrations ranging from 0.05 to 4 mg/ml. K_A values were calculated as described,[32] by determining the relative mobility of CBM29-2 and its derivatives in the presence and absence of ligand *versus* BSA (Sigma), the control non-binding protein. ITC measurements were made at 25 °C and 5 °C following standard procedures[11] using a Microcal Omega titration calorimeter. The proteins and ligand were in 50 mM sodium phosphate buffer (pH 7.5). During a titration experiment, the protein sample (100–800 μM), stirred at 300 rpm in a 1.4331 ml reaction cell maintained at 25 °C or 5 °C, was injected with 25 successive 10 μl aliquots of ligand comprising polysaccharide (20 mg/ml) or oligosaccharide (10 mM), at 200 s intervals. The molar concentration of CBM29 binding sites present on the polysaccharide ligand was determined as described.[12] Integrated heat effects, after correction for heats of dilution, were analyzed by non-linear regression using a single site-binding model (Microcal Origin, version 5.0). The fitted data yield the association constant (K_A) and the enthalpy of binding (ΔH). Other thermodynamic parameters were calculated using the standard thermodynamic equation:

$$-RT \ln K_A = \Delta G = \Delta H - T\Delta S$$

The *c*-values (product of the molar concentration of binding sites on the protein multiplied by the association constant) were predominantly ≥ 1, allowing accurate deconvolution of the binding data.

Analytical ultracentrifugation

Sedimentation equilibrium ultracentrifugation was performed in an Optima XL-A analytical ultracentrifuge (Beckman Instruments, Palo Alto, USA), using the absorption optical system. E78R was at a loading concentration of 350 μM in 50 mM sodium phosphate buffer (pH 7.5). Because of the high concentration of protein (necessary to probe for any self-association), a wavelength of 297 nm was chosen to reduce the absorption level into the Beer–Lambert range and short (4 mm) optical path-length double-sector cells were used. When appropriate, the ligand cellohexaose was added to a final concentration of 1 mM. Ultracentrifugation was carried out at 25.0 °C and run overnight at a rotor speed of 17,000 rpm (~22,600 *g*) until equilibrium distributions of solute in the cells was attained. A baseline was obtained by overspeeding.[33] Apparent weight average molecular masses M were determined from the equilibrium concentration distributions using the MSTAR procedure.[34,35] Estimates for the molar dissociation constants K_d ($= 1/K_a$, the molar association constant) and M were obtained using the program NONLIN.[36] MSTAR is a model-independent programme that calculates the weight average molecular mass over the whole distribution of macromolecules in the centrifuge cell: from meniscus to base. NONLIN, on the other hand, is a model-dependent programme that will find the molecular mass between a specified range of radial displacement data in the cell. This means that if the data are obscured near the cell base they will be excluded from the NONLIN analysis. MSTAR, however, obtains the missing data from a simple extrapolation. For a mixture, missing the data from near the cell base will lead to an in the molecular mass.

Crystallization and X-ray data collection

Purified CBM29-2 (E78R) at 10 mg/ml was crystallised in the presence of 5 mM cellohexaose in hanging-drops with 50 mM calcium chloride and 25% (w/v) PEG 3350. Prior to data collection, single crystals were harvested into a cryo-protectant mother liquor containing 25% PEG 3350, 15% (v/v) ethylene glycol and 100 mM calcium chloride, and frozen directly, in rayon fibre loops, into liquid N_2.

Data were collected, to 1.5 Å resolution, on beam-line ID14-EH1 at the European Synchrotron Radiation Facility (ESRF) in Grenoble (France). A total of 200 images of 0.5° per image were collected at 0.934 Å. Crystals belong to orthorhombic space group $P2_12_12_1$ ($a = 38.7$ Å, $b = 56.5$ Å, $c = 129.8$ Å) and have two molecules in the asymmetric unit. Data at a maximum resolution of 1.5 Å were processed and scaled using MOSFLM and SCALA from the CCP4 suite (R_{merge} 5.3%, completeness 93%, multiplicity 3.3, $I/\sigma I$ 16.0; see Table 3 for data and structural quality).[37] The structure was solved by molecular replacement using the program AMoRe[38] and the CBM29-2 cellohexaose complex (1GWM) as the search model. Following molecular replacement, maximum-likelihood-based refinement of the atomic positions and temperature factors was performed with REFMAC.[39] This procedure was interspersed with manual correction using the X-AUTOFIT option within QUANTA (Accelrys Inc. San Diego, CA). Water molecules were first placed automatically with REFMAC/ARP[39,40] and verified manually before coordinate deposition. The final model has R_{cryst} 0.17, R_{free} 0.19 and deviations from stereochemical target values of 0.014 Å for bonds and 1.5° for angles.

Database accession code

The coordinates for the structure described here have

Table 3. X-ray data and structural quality statistics for the E78R mutant of CBM29-2

Radiation source	ID14-EH1
No. images	200
Detector	ADSC quantum4
Software	Mosflm/Scala
Space group	$P2_12_12_1$
Unit cell parameters	
a (Å)	38.7
b (Å)	56.5
c (Å)	129.8
Wavelength (Å)	0.934
Resolution (Å)	43.0–1.5 (1.58–1.50)
Unique reflections	42,994
Completeness (%)	92.6 (68.4)
R_{sym} (%)	5.3 (47.1
Multiplicity	3.3 (1.9)
$I/\sigma I$	16.0 (1.8)
Refinement	
R_{cryst} (%)	17.0
R_{free} (%)	18.7
RMS on bond distances (Å)	0.014
RMS on bond angles (deg.)	1.488

been deposited with the Macromolecular Structures Database with accession code 1oh3.

References

1. Gill, J., Rixon, J. E., Bolam, D. N., McQueen-Mason, S., Simpson, P. J., Williamson, M. P. *et al.* (1999). The type II and X cellulose-binding domains of Pseudomonas xylanase A potentiate catalytic activity against complex substrates by a common mechanism. *Biochem. J.* **342**, 473–480.
2. Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E. *et al.* (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem. J.* **331**, 775–781.
3. Coutinho, P. M. & Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G. J., Henrissat, B. & Svensson, B., eds), pp. 3–12, The Royal Society of Chemistry, Cambridge.
4. Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J. & Williamson, M. P. (2000). The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *J. Biol. Chem.* **275**, 41137–41142.
5. Boraston, A. B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D. R., Kilburn, D. G. & Davies, G. J. (2002). Differential oligosaccharide recognition by evolutionarily-related beta-1,4 and beta-1,3 glucan-binding modules. *J. Mol. Biol.* **319**, 1143–1156.
6. Czjzek, M., Bolam, D. N., Mosbah, A., Allouch, J., Fontes, C. M., Ferreira, L. M. *et al.* (2001). The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. *J. Biol. Chem.* **276**, 48580–48587.
7. Simpson, P. J., Jamieson, S. J., Abou-Hachem, M., Karlsson, E. N., Gilbert, H. J., Holst, O. & Williamson, M. P. (2002). The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase. *Biochemistry,* **41**, 5712–5719.
8. Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M. & Teeri, T. T. (2003). The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc. Natl Acad. Sci. USA*, **100**, 484–489.
9. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. & Steitz, T. A. (1996). Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J.* **15**, 5739–5751.
10. Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J. & Fontes, C. M. (2000). The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain. *Biochemistry,* **39**, 5013–5021.
11. Szabo, L., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2001). Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. evidence that xylan binds in an approximate 3-fold helical conformation. *J. Biol. Chem.* **276**, 49061–49065.
12. Charnock, S. J., Bolam, D. N., Nurizzo, D., Szabo, L., McKie, V. A., Gilbert, H. J. & Davies, G. J. (2002). Promiscuity in ligand-binding: the three-dimensional structure of a *Piromyces* carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. *Proc. Natl Acad. Sci. USA*, **99**, 14077–14082.
13. Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A. *et al.* (2001). Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. *Biochemistry,* **40**, 6240–6247.
14. Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G. & Rose, D. R. (2002). High-resolution crystal structures of the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry,* **41**, 4246–4254.
15. Xie, H., Bolam, D. N., Nagy, T., Szabo, L., Cooper, A., Simpson, P. J. *et al.* (2001). Role of hydrogen bonding in the interaction between a xylan binding module and xylan. *Biochemistry,* **40**, 5700–5707.
16. Xie, H., Gilbert, H. J., Charnock, S. J., Davies, G. J., Williamson, M. P., Simpson, P. J. *et al.* (2001). *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. *Biochemistry,* **40**, 9167–9176.
17. Nagy, T., Simpson, P., Williamson, M. P., Hazlewood, G. P., Gilbert, H. J. & Orosz, L. (1998). All three surface tryptophans in type IIa cellulose binding domains play a pivotal role in binding both soluble and insoluble ligands. *FEBS Letters*, **429**, 312–316.
18. McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A. & Kilburn, D. G. (2000). Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. *Protein Eng.* **13**, 801–809.
19. Kormos, J., Johnson, P. E., Brun, E., Tomme, P., McIntosh, L. P., Haynes, C. A. & Kilburn, D. G. (2000). Binding site analysis of cellulose binding domain CBD(N1) from endoglucanse C of *Cellulomonas fimi* by site-directed mutagenesis. *Biochemistry,* **39**, 8844–8852.
20. Linder, M., Lindeberg, G., Reinikainen, T., Teeri, T. T. & Pettersson, G. (1995). The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution. *FEBS Letters*, **372**, 96–98.
21. Simpson, H. D. & Barras, F. (1999). Functional analysis of the carbohydrate-binding domains of *Erwinia chrysanthemi* Cel5 (endoglucanase Z) and an *Escherichia coli* putative chitinase. *J. Bacteriol.* **181**, 4611–4616.
22. Brett, C. T. & Waldren, K. (1996). *Physiology and Biochemistry of Plant Cell Walls. Topics in Plant Functional Biology* (Black, M. & Charlewood, B., eds), vol. 1, Chapman & Hall, London.
23. Freelove, A. C., Bolam, D. N., White, P., Hazlewood, G. P. & Gilbert, H. J. (2001). A novel carbohydrate-binding protein is a component of the plant cell wall-degrading complex of *Piromyces equi*. *J. Biol. Chem.* **276**, 43010–43017.
24. Lee, S. S., Ha, J. K. & Cheng, K. (2000). Relative contributions of bacteria, protozoa, and fungi to *in vitro* degradation of orchard grass cell walls and their interactions. *Appl. Environ. Microbiol.* **66**, 3807–3813.
25. Berg, J. M., Tymoczko, J. L. & Stryer, L. (2001). Prelude: biochemistry and the genomic revolution. In *Biochemistry,* 5th edit., W. H. Freeman and Co, New York p. 11.
26. Linder, M., Mattinen, M. L., Kontteli, M., Lindeberg,

G., Stahlberg, J., Drakenberg, T. *et al.* (1995). Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. *Protein Sci.* **4**, 1056–1064.
27. Pell, G., Williamson, M. P., Walters, C., Du, H., Gilbert, H. J. & Bolam, D. N. (2004). Importance of hydrophobic and polar residues in ligand binding in the family is carbohydrate-binding module from *Cellvibrion japonicus* Xynloc. *Biochemistry*, **42**, 9316–9323.
28. Notenboom, V., Boraston, A. B., Chiu, P., Freelove, A. C., Kilburn, D. G. & Rose, D. R. (2001). Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. *J. Mol. Biol.* **314**, 797–806.
29. Lis, H. & Sharon, N. (1998). Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637–674.
30. Wright, C. S. (1990). 2.2 A resolution structure analysis of two refined *N*-acetylneuraminyl-lactose–wheat germ agglutinin isolectin complexes. *J. Mol. Biol.* **215**, 635–651.
31. Boraston, A. B., Revett, T. J., Boraston, C. M., Nurizzo, D. & Davies, G. J. (2003). Structural and thermodynamic dissection of specific mannan recognition by a carbohydrate binding module, TmCBM27. *Structure*, **11**, 665–675.
32. Takeo, K. (1984). Affinity electrophoresis-principles and applications. *Electrophoresis*, **5**, 187–195.
33. Winzor, D. J. & Harding, S. E. (2001). Sedimentation equilibrium in the analytical ultracentrifuge. In *Protein–Ligand Interactions: Hydrodynamics and Calorimetry* (Harding, S. E. & Chowdhry, B. Z., eds), pp. 105–135, Oxford University Press, Oxford.
34. Cölfen, H. & Harding, S. E. (1997). MSTARA and MSTARI: interactive PC algorithms for simple, model independent evaluation of sedimentation equilibrium data. *Eur. Biophys. J.* **25**, 333–346.
35. Creeth, J. M. & Harding, S. E. (1982). Some observations on a new type of point average molecular weight. *J. Biochem. Biophys. Methods*, **7**, 25–34.
36. Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halvorson, H. R. (1981). Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* **36**, 575–588.
37. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
38. Navaza, J. & Saludijan, P. (1997). AMoRe: an automated molecular replacement program package. *Methods Enzymol.* **276**, 581–594.
39. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
40. Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallog. sect. D*, **49**, 129–147.
41. Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–134.

Edited by J. Thornton

(Received 6 November 2003; received in revised form 18 December 2003; accepted 22 December 2003)