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[†,‡]

Valerie Notenboom,*[,§,||,⊥] Alisdair B. Boraston,[§,⊥,@,#] Spencer J. Williams,[§,+] Douglas G. Kilburn,[§,@] and David R. Rose[§,||]

Protein Engineering Networks of Centres of Excellence and Department of Microbiology and Immunology and Biotechnology Laboratory, University of British Columbia, Vancouver, BC, Canada, and Department of Medical Biophysics and Ontario Cancer Institute, University of Toronto, Toronto, ON, Canada

Received October 17, 2001; Revised Manuscript Received January 18, 2002

ABSTRACT: Carbohydrate-binding module (CBM) family 13 includes the "R-type" or "ricin superfamily" β-trefoil lectins. The C-terminal CBM, CBM13, of xylanase 10A from *Streptomyces lividans* is a family 13 CBM that is not only structurally similar to the "R-type" lectins but also somewhat functionally similar. The primary function of CBM13 is to bind the polysaccharide xylan, but it retains the ability of the R-type lectins to bind small sugars such as lactose and galactose. The association of CBM13 with xylan appears to involve cooperative and additive participation of three binding pockets in each of the three trefoil domains of CBM13, suggesting a novel mechanism of CBM−xylan interaction. Thus, the interaction of CBM13 with sugars displays considerable plasticity for which we provide a structural rationale. The high-resolution crystal structure of CBM13 was determined by multiple anomalous dispersion from a complex of CBM13 with a brominated ligand. Crystal structures of CBM13 in complex with lactose and xylopentaose revealed two distinct mechanisms of ligand binding. CBM13 has retained its specificity for lactose via Ricin-like binding in all of the three classic trefoil binding pockets. However, CBM13 has the ability to bind either the nonreducing galactosyl moiety or the reducing glucosyl moiety of lactose. The mode of xylopentaose binding suggests adaptive mutations in the trefoil sugar binding scaffold to accommodate internal binding on helical polymers of xylose.

Most glycoside hydrolases originating from polysaccharolytic organisms are modular proteins comprising a catalytic domain and one or more carbohydrate-binding modules (CBMs),[1] which are usually, but not exclusively, complementary in specificity to their catalytic component. CBMs are thought to function mainly by localizing the enzyme to the target substrate and, in some cases, by disrupting the structural integrity of some of the more robust polymers such as cellulose. To date, CBMs have been classified into 29 families based on sequence and structural similarities (*1*, *2*), and the crystal or NMR structures of representatives of more than 15 families have been determined.

The "ricin superfamily", or "R-type" lectin family, is a group of modules containing the hallmarks of the β-trefoil fold. The most notable examples from this group are galactose specific plant lectins (e.g., from *Ricinus communis*, *Abris precatorius*, *Sambucus nigra*, and *Viscum album*); however, examples can also be found in animals and microbes. They are found as modules in diverse classes of proteins such as mammalian glycosyl transferases, bacterial toxins, bacterial and fungal hydrolases, an arthropod serine protease, and a psychrophilic bacterial pectate lyase. Several species of bacteria, most notably the streptomycetes, have employed ricin superfamily lectins as the CBM in some of their glycoside hydrolases. This family of modules has subsequently been classified as family 13 CBMs (*1*).

These β-trefoil fold lectins are thought to have arisen from an early gene fusion event involving three repeats of a 42-amino acid peptide (*3*). Each of these peptides contains a disulfide bridge and two β-hairpin turns comprising a four-stranded trefoil peptide that is the scaffold for a single carbohydrate-binding site. Through mutation of binding site residues, many modules in the plant lectin members of this family have lost the ability to bind sugars in one or more of its three binding sites or have developed specificity for a particular carbohydrate. However, the overall β-trefoil fold remains. This fold, therefore, appears to be a scaffold into which an appropriate carbohydrate binding function can evolve through simple mutations.

[†] S.J.W. thanks the Killam Trust for a postdoctoral fellowship.

[‡] Coordinates for the structures described in this paper are accessible as PDB entries 1KNL, 1KNM, and 1KNN.

* To whom correspondence should be addressed. Current address: Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. E-mail: valerie@nki.nl. Phone: +31-20-5121945. Fax: +31-20-5121954.

[§] Protein Engineering Networks of Centres of Excellence.

[||] University of Toronto.

[⊥] These authors contributed equally to this paper.

[@] Department of Microbiology and Immunology and the Biotechnology Laboratory, University of British Columbia.

[#] Current address: Department of Chemistry, University of York, Heslington, York YO10 5DD, U.K.

[+] Current address: Department of Chemistry, University of California, Berkeley, CA 94720-1460.

[1] Abbreviations: CBM, carbohydrate binding module; MAD, multiple anomalous dispersion; CBr, 4-phenylbromide cellobiose.

10.1021/bi015865j CCC: $22.00
Published on Web 03/09/2002

Streptomyces lividans xylanase 10A contains an N-terminal family 10 endoxylanase and a C-terminal 14 kDa CBM belonging to family 13 (CBM13). CBM13 binds soluble and insoluble xylan, a polymer of xylose, as well as a variety of small soluble sugars, including galactose, lactose, xylooligosaccharides, and arabinooligosaccharides (*4*). CBM13 is one of the few sequences in this family of CBMs that has retained all three of its original β-trefoil sugar binding motifs and disulfide bridges, suggesting all three pockets can still individually bind small soluble saccharides. In fact, each of these binding sites has been shown to be important in binding xylan through an as yet unknown binding mechanism involving cooperativity between at least two of the binding sites.

This study investigates the recognition of xylan by CBM13 by X-ray crystallography and discusses its evolutionary relationship with other β-trefoil carbohydrate-binding proteins. Along with a complementary NMR spectroscopic study reported in this issue (*5*), the structures of several complexes, including a xylopentaose complex, reveal a structural rationale for recognition of xylan by CBM13, while maintaining its classic ability to bind small soluble sugars. We postulate that the original β-trefoil scaffold can be adapted to accommodate a helical xylan polymer by only a few permutations in nonconserved regions.

MATERIALS AND METHODS

Protein Production and Purification. CBM13 was produced and purified as described previously (*4*).

Synthesis of 4-Bromophenyl β-Cellobioside. 4-Bromophenyl β-cellobioside was synthesized according to the procedure of Ziser et al. (*6*). Briefly, a suspension of powdered 3 Å molecular sieves, silver carbonate, and 2,6-lutidine in dry acetonitrile was stirred for 5 min before the addition of hepta-*O*-acetyl-α-cellobiosyl bromide (*7*). The mixture was stirred at room temperature overnight. After workup, the compound was purified by flash chromatography to afford 4-bromophenyl hepta-*O*-acetyl-β-cellobioside. This material was deacetylated using sodium methoxide in methanol and crystallized from methanol and diethyl ether to afford 4-bromophenyl β-cellobioside as a colorless powder. The ^{1}H NMR data of this material were in close agreement with those reported previously (*6*).

Crystallization and Data Collection. Crystals of CBM13 appeared overnight from 1.8 M ammonium sulfate, 50 mM MES (pH 6.5), and 5% dioxane in hanging or sitting drops, from a protein concentration of 30 mg/mL. Native crystals were of the tetragonal space group $P4_12_12$ (allowing for one molecule in the asymmetric unit) with the following cell dimensions: $a = b = 55.95$ Å and $c = 97.07$ Å. As CBM13 contains no methionine residues and the native data set resisted being solved by molecular replacement with related structures available at the time in the RCSB protein database, a brominated substrate was synthesized to add an ordered anomalous scatterer. Crystals of CBM13 complexed with 4-phenylbromide cellobiose (CBr) grew from similar conditions to native CBM13 crystals, but were orthorhombic in space group $P2_12_12_1$ ($a = 32.57$ Å, $b = 58.99$ Å, and $c = 116.37$ Å) with two molecules per asymmetric unit. A three-wavelength MAD data set was collected to 1.5 Å resolution at the Advanced Photon Source (APS) at Argonne National Laboratories (Argonne, IL) at the BioCARS 14-BM-D, on a Q4 CCD area detector. The absorption edge for the Br atom was detected by scanning through the theoretical X-ray absorption edge. From this, three wavelengths were chosen, one at the inflection point or edge of the absorption profile (f', $\lambda = 0.91994$ Å), one at its peak (f'', $\lambda = 0.91976$ Å), and one reference set at a remote wavelength ($\lambda = 0.89844$ Å). The program package Denzo/Scalepack was used to refine experimental parameters and process reflections (*8*). Crystals of CBM13 in the presence of excess lactose and xylopentaose were also grown from the same conditions, but again yielded crystals with alternative symmetry. The CBM13−lactose complex crystallized in space group $P2_12_12_1$ and diffracted to 1.2 Å resolution (one molecule in the asymmetric unit), with the following unit cell dimensions: $a = 43.19$ Å, $b = 46.20$ Å, and $c = 56.46$ Å. The CBM13−xylopentaose complex formed very small trigonal $P3_221$ crystals (30 μm × 30 μm × 150 μm) (one molecule per asymmetric unit) with the following unit cell lengths: $a = b = 55.13$ Å and $c = 80.14$ Å. These crystals diffracted to 1.7 Å resolution. All of the nonanomalous data were collected at APS at the BioCARS 14-BM-C station, on a Quantum 4 CCD area detector at an X-ray wavelength of 1.0 Å, and scaled with the Denzo/Scalepack package (*8*). Data collection statistics are listed in Table 1.

Table 1: Data Collection Statistics

	edge (Br)	peak (Br)	remote (Br)	native
wavelength (Å)	0.91944	0.91976	0.89844	1.0000
space group	$P2_12_12_1$			$P4_12_12$
maximum resolution (Å)	1.5	1.5	1.5	1.0
no. of observations	197157	199481	208350	573332
no. of unique reflections	34575	34721	35531	65601
I/σ	14.9 (11.2)	15.3 (11.4)	16.8 (10.8)	26.9 (15.3)
completeness (%)	93.4 (58.7)	94.0 (61.5)	96.2 (73.6)	89.5 (78.1)
R_{merge} (%)	5.2 (15.4)	5.0 (14.5)	4.7 (12.8)	5.7 (26.9)

	lactose	xylopentaose
wavelength (Å)	1.0000	1.0000
space group	$P2_12_12_1$	$P3_221$
maximum resolution (Å)	1.2	1.7
no. of observations	126516	108110
no. of unique reflections	30016	13471
I/σ	18.7 (9.3)	11.4 (5.6)
completeness (%)	83.1 (22.1)	83.1 (23.4)
R_{merge} (%)	4.1 (19.3)	8.0 (29.3)

Structure Determination and Refinement. The program Solve was used to locate both Br atoms in the asymmetric unit (20 and 42% occupancy), and experimental phases were determined (*9*). The mean figure of merit was 0.34 over all data, and 0.79 after 10 cycles of density modification using DM, implementing solvent flattening, and histogram matching (*10*, *11*). Despite the low Br atom occupancies, the resulting 1.5 Å experimental map was of superb quality, with unambiguous and continuous electron density throughout the asymmetric unit, including two CBr molecules (Figure 1). Both protein chains were manually traced and fitted in O (*12*). The two resulting molecules underwent rigid body refinement and simulated annealing in the program CNS (*13*), implementing bulk solvent correction and a maximum likelihood refinement target using structure factor amplitudes, against the data collected at the remote wavelength. A random 7% (6332) of reflections were set aside for cross-validation purposes (*14*).

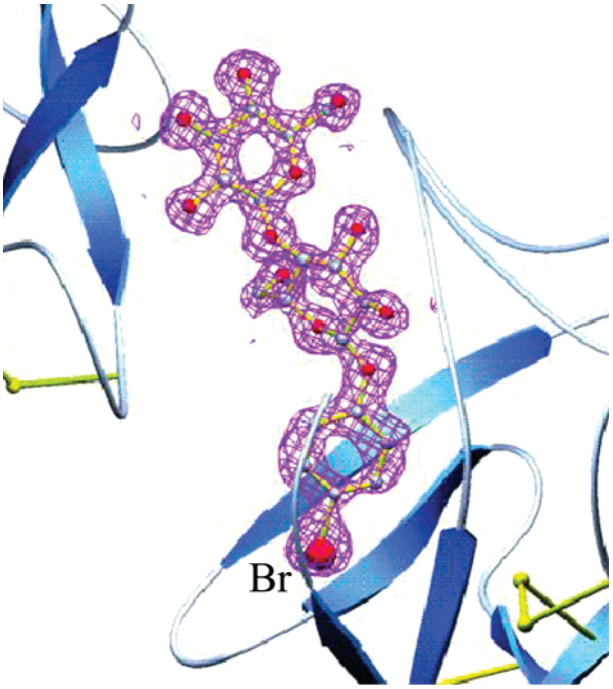

FIGURE 1: Solvent-flattened electron density to 1.5 Å resolution around substrate atoms calculated from experimental Br-MAD phases. This figure was created with Setor (*32*, *33*).

This partially refined structure of CBM13-CBr was used in the program AMoRe (*15*) to find the correct orientation of CBM13 in the asymmetric unit of the tetragonal crystals of the unliganded protein, the orthorhombic lactose complex, and the trigonal xylooligomer complex crystals. In all cases, the solution was readily found without ambiguity. This model was used for complete automatic rebuilding and iterative refinement using the program ARP/wARP (*16*). The appropriate saccharide was then fit into the resulting difference density, and the solvent was built with ARP/wARP (*16*). Refinement was completed using Refmac5 within the program package CCP4 (*11*, *17*), with a maximum likelihood target using structure factor amplitudes, refining anisotropic *B*-factors.

Refinement and structure quality statistics are listed in Table 2.

RESULTS

Description of the General Structure of CBM13. CBM13 assumes a β-trefoil fold displaying the characteristic pseudo-3-fold symmetry arising from tandem α, β, and γ peptide repeats, in accordance with predictions based on the similarity of the sequence to the ricin B chain (RTB) (*4*), as well as an analysis of NMR chemical shift and ^{15}N relaxation data (*5*). The three subdomains form a β-barrel fold containing 12 sheets, with five conserved residues contributed by each of the three peptide repeats forming a compact hydrophobic core (Figure 2). Each subdomain also contains a disulfide bridge, a feature conserved from the original trefoil domain that stabilizes the peptide fold. The native structure of CBM13 was of sufficient resolution to determine that in two of the three putative binding pockets an ordered glycerol molecule was bound (not shown). The short soak in a 30% glycerol cryoprotection solution prior to data collection is the likely source of these molecules, considering that none of the purification buffers contained glycerol.

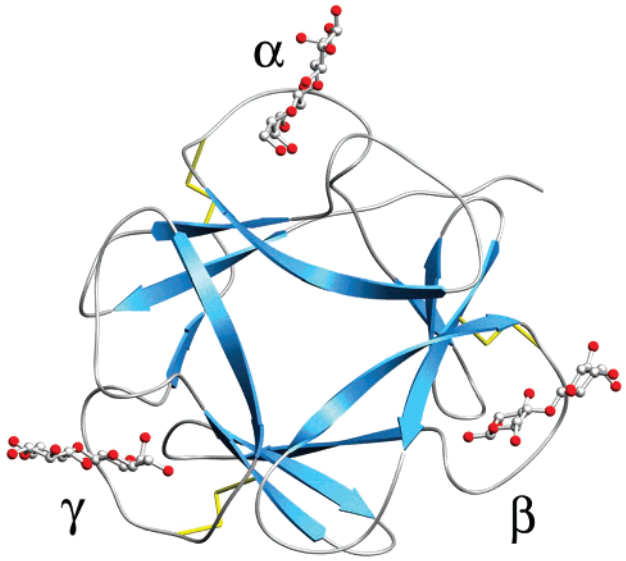

FIGURE 2: Ribbon diagram of the *S. lividans* CBM13 crystal structure with lactose moieties bound in all three subsites of the trefoil fold. This figure was created with Setor (*32*, *33*).

Table 2: Structure Quality Statistics

	native (to 1.2 Å)	lactose (1.2 Å)	xylopentaose (1.7 Å)
R-factor, all data (%)	13.9 (19.7)	12.3 (23.0)	16.4 (33.1)
R_{free}, all data (%)	16.6 (22.9)	16.9 (23.4)	22.0 (33.2)
rms deviation from ideal			
bond lengths (Å)	0.008	0.008	0.007
bond angles (deg)	1.22	1.37	1.44
average *B*-factor (Å^2)	18.5	14.6	12.8
	Substrate Atoms		
α	Gal, 12.6 Glc, 40.7	37.8, 11.2, 10.6, 17.8, 32.2	
β (shared with γ)	Glc, 7.7	(37.8, 11.2, 10.6, 17.8, 32.2)	
γ (shared with β)	Gal, 5.9		

CBM13 with Lactose. CBM13 in complex with lactose shows substrate bound in all three trefoil binding pockets. The α pocket (Figure 3a) binds lactose at its nonreducing galactose moiety through interactions with all of its conserved residues (Asp22, Gln32, Asn41, and Gln42) and a stacking interaction with Trp34. The glucose moiety at the reducing end of the lactose showed very little density and was presumed to be disordered. The γ pocket also binds galactose with similar interactions in the binding pocket with all of the expected residues (Figure 3b). Tyr117 stacks with the pyranose ring of galactose, whereas Asp102, Gln115, Asn124, and Gln125 form hydrogen bonding interactions. The glucose unit of the lactose molecule in the γ pocket is ordered and bound by the β site of a crystallographically related CBM13 molecule (not shown). Because it also makes classic interactions with the ligand, combined with the previous observation that all three binding pockets are capable of binding sugars (*4*), this is presumed not to be an artifact brought on by crystallization but indeed a relevant interaction.

CBM13 in Complex with Xylopentaose. The CBM13−xylopentaose complex structure shows clear density for all five xylose units of the oligosaccharide (Figure 4). Xylopentaose occupies both the α and β binding sites of CBM13 simultaneously, lying across the pockets of two successive CBM13 molecules (Figure 4b), forming crystal contacts

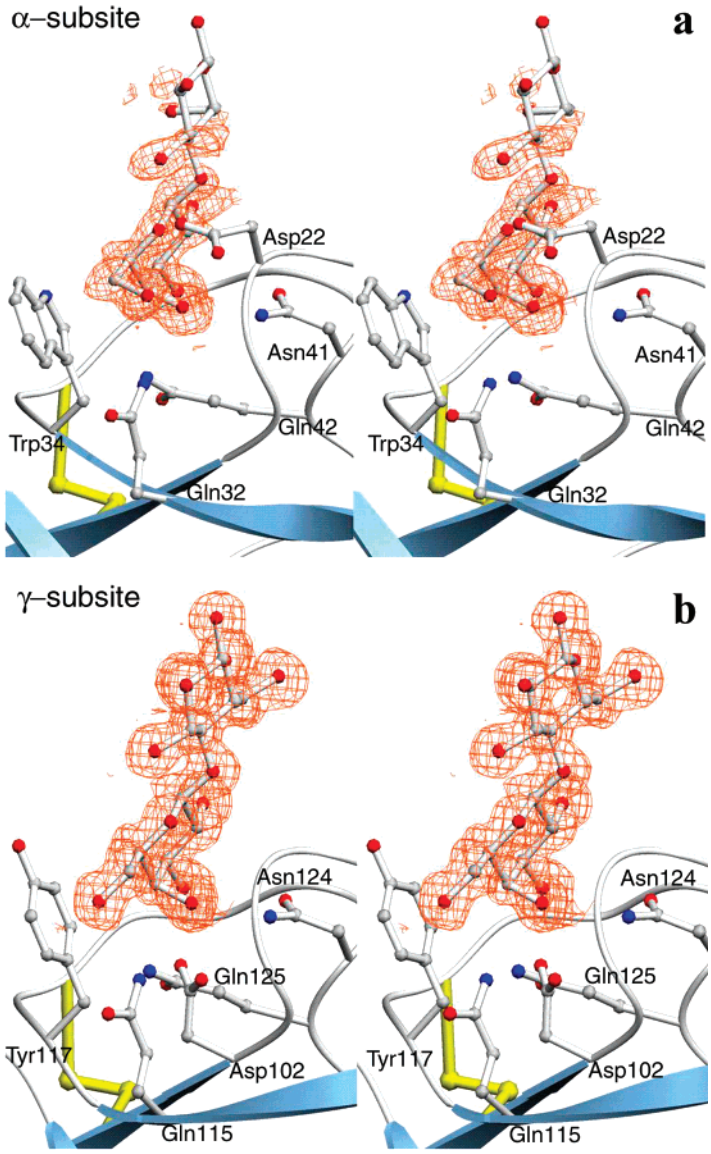

FIGURE 3: (a and b) Stereorepresentation of $F_o - F_c$ difference density at 1.2 Å resolution in the α and γ binding pockets of CBM13 cocrystallized with lactose. Final refined substrate atoms have been superimposed for clarity. This figure was created with Setor (*32*, *33*).

between asymmetric units. The α pocket mainly interacts with the second xylose unit from the reducing end, whereas the main interactions in the β pocket are with the third or middle xylose unit of the ligand. Because oxygen and carbon atoms cannot be distinguished from each other from electron density alone at this resolution, the distinction between the reducing and nonreducing end of the xylooligomer could not be made. Hydrogen bonding patterns and *B*-factor analysis adequately resolved this ambiguity, where the final direction of the sugar corresponded to that of the CBr ligand in the β pocket.

The conserved residues in the α and β pockets form the bulk of the interactions with xylopentaose. In addition, two nonconserved residues at the top of the sugar binding β-hairpin loops, His37 and Trp77 in the α and β pockets, respectively (shown in orange in Figures 4b), form interactions with adjacent xylose units. Data from the following paper (*5*) show that, although xylopentaose binds to all three sites in solution, binding to the β pocket is significantly favored over binding to the other sites and suggests that the additional hydrophobic interactions with Trp77 that cannot be provided by a His in the corresponding position might play a role in defining that difference.

DISCUSSION

Mechanisms of Xylan Recognition. Xylan binding activity has been found for examples of CBMs from six of the 29 CBM families. These are the family 2b CBMs from *Cellulomonas fimi* Xyn11A (CfCBM2b-1 and CfCBM2b-2), the family 4 CBMs from *Rhodothermus marinus* Xyn10A, the family 6 CBMs from *Clostridium stercorarium* Xyn10A, the family 9b CBM from *Thermotoga maritima* Xyn10A (TmCBM9b-2), the family 13 CBMs from *S. lividans* Xyn10A (CBM13) and *S. olivaceoviridis* Xyn10A, and family 22 CBMs from several bacterial species, including CtCBM22-2 from *Clostridium thermocellum* Xyn10A. Of these, detailed biochemical and structural studies have been done on CfCBM2b-1 (*18*), CfCBM2b-2 (*19*), TmCBM9b-2 (*20*, *21*), and CtCBM22-2 (*22*).

CfCBM2b-1 and CtCBM22-2 bind xylooligomers at extended binding sites that accommodate sugars four to five

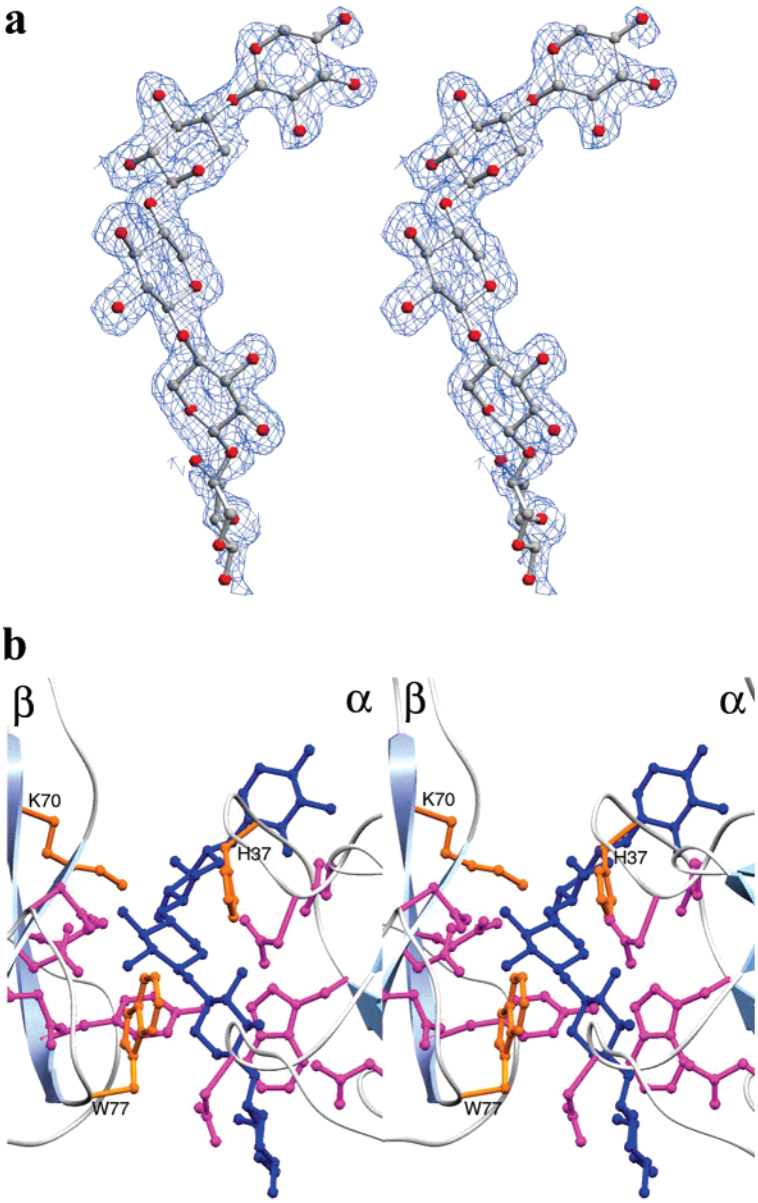

FIGURE 4: (a and b) Stereorepresentation of (a) $F_o - F_c$ difference density at 1.7 Å of CBM13 cocrystallized with xylopentaose with final refined substrate atoms superimposed for clarity. In panel b, the α and β binding pockets share one molecule of xylopentaose (blue), forming a crystal contact. Residues depicted in magenta are those conserved in the trefoil peptide motif, and those colored in orange are nonconserved residues that form interactions with the xylooligomer. This figure was created with Setor (*32*, *33*).

monosaccharide units in length. Both binding sites contain two or more aromatic residues that are thought to interact with pyranoside rings. However, while the binding site of CfCBM2b-1 is shallow and formed mainly by the perpendicular orientation of two surface tryptophan residues, the binding site of CtCBM22-2 is somewhat deeper and formed by two loops containing aromatic residues that face one another such that they can sandwich and nearly sequester the sugar chain. The shape of the CfCBM2b-1 binding site is thought to complement the helical shape of xylan. It is unclear whether this is true for the binding site of CtCBM22-2, although the interaction with an extended ligand suggests that this is likely. Both CfCBM2b-1 and CtCBM22-2 can bind internally on a xylan chain.

Though the architecture of the TmCBM9b-2 binding site is similar to that of CtCBM22-2 with tryptophan residues that sandwich the ligand, one end of the binding site is blocked such that it can only bind the reducing ends of xylan chains. Only a disaccharide, e.g., xylobiose, is required to satisfy the complete complement of protein−carbohydrate interactions with TmCBM9b-2. There is no indication that the helical structure of xylan plays any role in its recognition by TmCBM9b-2.

The architecture of the α and β binding sites of CBM13 with respect to xylooligosaccharide recognition is ostensibly similar to that of CfCBM2b-1 and CfCBM2b-2. The NMR data show enhanced binding of xylooligomers up to xylotetraose, with little improvement beyond (*5*). Indeed, the binding sites are extended, accommodating a sugar three to four units in length, and have hydrogen bonding polar residues running the length of the site. They also contain two residues that provide specific sugar interactions (tryptophans, tyrosines, or histidines) positioned approximately 5−7 Å apart. Like those of CfCBM2b-1 and CfCBM2b-2,

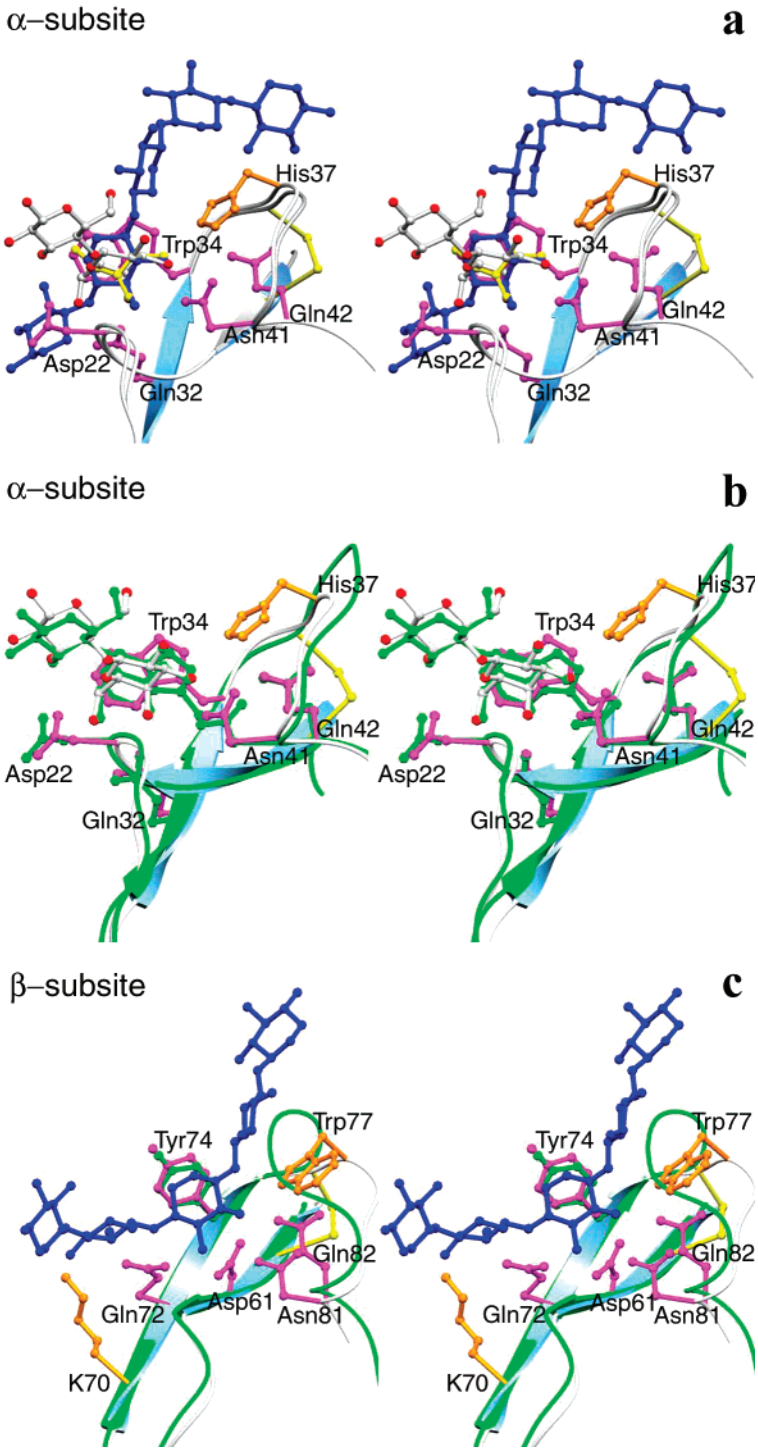

FIGURE 5: (a−c) Residues depicted in magenta are those conserved in the trefoil peptide motif, and those colored in orange are nonconserved residues that form interactions with the sugars. Stereorepresentation of (a) the α binding pocket with lactose (white bonds), glycerol (yellow), and xylopentaose (blue) bound. Lactose is recognized in a perpendicular fashion, whereas xylopentaose lies across the binding site, maintaining a helical structure. (b) Superposition of the α binding pockets of the CBM13−lactose complex (white bonds) and the ricin B chain−lactose complex [green, with conserved residues (PDB entry 2AAI)]. (c) Superposition of the *β* binding pockets of the CBM13−xylopentaose complex (white bonds and blue substrate) and the ricin B chain (green, no substrate bound). This figure was created with Setor (*32*, *33*).

the shape of the α and *β* binding sites of CBM13 mirrors the helical structure observed in the bound xylooligosaccharides (Figure 5a,c). This mode of binding, which does not engage the ends of the oligosaccharide, would allow CBM13 to bind internally on a xylan chain, as suggested by the crystal packing of the xylopentaose complex (Figure 6). The binding sites of CBM13 interact mainly with three xylose units of a xylooligosaccharide. The central unit fits in what is considered the binding pocket in RTB. The 2′- and 3′-OH groups of this sugar unit are directed into the binding pocket; thus, arabinosyl or glucuronyl side chains could not be accommodated at these positions on this sugar unit. However, the 2′- and 3′-OH groups of the flanking sugar units are directed into the solvent so side chain decorations may be accommodated on these sugar units. CBM13 does bind to glucuronic acid (A. Boraston, unpublished results),

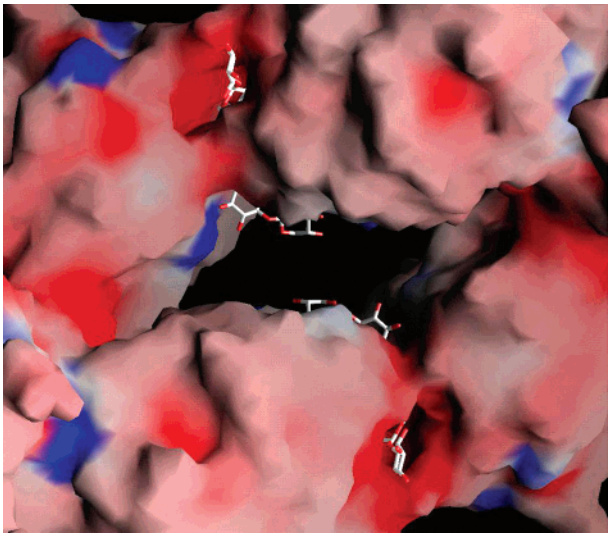

FIGURE 6: Surface representation of crystal packing of CBM13 $P3_221$ crystals in complex with xylopentaose. The α and β pockets form a tunnel recognizing the helical structure of a xylan oligomer.

and the structures presented here provide no reason to exclude the possible recognition of glucuronic acid side chains as found in Birchwood xylan.

The binding sites in the α, β, and γ pockets appear to cooperate in the binding of CBM13 to xylan (*4*). We hypothesized that this resulted from the xylan polymer spanning two of the binding sites on the same module. However, upon examination of the conformation and orientation of xylopentaose in the CBM13−xylopentaose complex structure, it appears that longer xylose polymers could only span two sites if a considerable bend was introduced into the chain, an energetically highly unfavorable situation. The CBM13−xylopentaose complex had single xylopentaose molecules occupying the α site of one module and the β site of an adjacent module. This suggests that single binding sites of two CBM13 molecules have the capacity to interact simultaneously with the same ligand in solution. However, NMR experiments did not suggest the presence of any significant protein−protein interactions in solution (*5*). Conversely, as discussed in more detail in the following paper (*5*), the crystallographic observations suggest the ability of an individual CBM to bind concurrently to more than one strand of oligomeric xylan. Thus, it appears that the cooperativity between the binding sites of CBM13 occurs through its multiple binding sites interacting with separate xylan chains. This situation requires that the xylan chains be immobilized (i.e., have fixed positions relative to one another) such as might be expected in insoluble xylan or affinity electrophoresis where the xylan polymers are imbedded in the gel matrix.

The β-Trefoil Fold, a General Scaffold for Carbohydrate Recognition. The β-trefoil fold was first observed in the structure of soybean trypsin inhibitor (*23*). It has subsequently been found in proteins of diverse functions. The fold contains 12 strands of β-sheet, forming six hairpin turns. A β-barrel structure is formed by six of the strands, in conjunction with three hairpin turns. The other three hairpin turns form a triangular cap on one end of the β-barrel called the "hairpin triplet". The subunit of this fold, called here a trefoil domain, is a contiguous amino acid sequence with a four-β-strand, two-hairpin structure having a trefoil shape. Each trefoil domain contributes one hairpin (two β-strands) to the β-barrel and one hairpin to the hairpin triplet. The fold of the resulting molecule has a pseudo-3-fold axis (*24, 25*).

The primary structures of carbohydrate binding polypeptides with known β-trefoil folds can be dissimilar with levels of identity ranging from 15 to 90%. The dissimilarity occurs primarily in the loop regions. The hydrophobic amino acid side chains that form the hydrophobic contacts in the barrel are well-conserved in polypeptides adopting this fold, not just those that bind carbohydrates (*24*) (Figure 7). Disulfide bonds occurring within the trefoil domains are also common. Most notably, CBM13 and SoCBM13 (*26*), which are 90% identical in amino acid sequence and have an rmsd of 0.43 $Å^2$ over 484 backbone atoms, have retained their disulfide bonds in the α, β, and γ domains, while in the plant lectins, these bonds still appear in the α and β domains.

In the family 13 CBMs, the substrate binding sites are small pockets formed between two hairpins of a single trefoil domain. The binding residues, which are remarkably conserved between the plant β-trefoil lectins and CBM13, occur mainly in the β-sheets or immediately adjacent to β-sheets (Figure 7) and overlap well in known structures. Though the human cysteine-rich macrophage mannose receptor shares only one conserved binding residue in the γ site (*31*), namely, an aspartic acid, the carbohydrate recognition site resides in the same region as for CBM13 and the plant lectins.

Though the β-trefoil fold is often termed a galactose binding fold (*27*) and is thought to originate from the triplication of an ancestral 42-amino acid galactose binding polypeptide (*3*), the trefoil domains of this fold can have carbohydrate binding sites with very different specificities. The 1α and 2γ domains of RTB both bind lactose and galactose, but only the 2γ site binds *N*-acetylgalactosamine. Other less well characterized bacterial family 13 CBMs bind to mannose (*28*) or β-1,3-glucan (*29*). The human cysteine-rich macrophage mannose receptor has the most profoundly different binding specificity; it binds chondroitin sulfate and blood group Lewis(a) and Lewis(x) carbohydrates, i.e., sulfated oligosaccharides (*30*). Furthermore, it does so via a nonconserved complement of binding residues (*31*) (Figure 7).

The binding sites of CBM13 can accommodate a variety of mono-, di-, and oligosaccharides, including galactose and lactose (*4*). As shown here, CBM13 binds lactose in a manner identical to that of RTB with the nonreducing end pointing into the base of the binding site (Figure 5b). However, CBM13 has acquired the ability to bind polymers of xylose, which are in fact the preferred ligands (*4, 5*), such that the reducing and nonreducing ends are free in solution (Figure 4). This appears to be conferred, at least in part, by the introduction into the binding sites of amino acids with side chains that are capable of interacting with the sugar. This is most evident in the β site of CBM13 where Trp77, a nonconserved residue (Figure 7), interacts with a pyranose ring of the bound xylooligosaccharide (Figure 5c). At the equivalent position in the α site, His37 is positioned to form a hydrogen bond with the substrate (Figure 4a). The presence of these aromatic side chains and their perpendicular orientation relative to the conserved binding site makes the α and β binding sites extended and able to specifically accommodate the helical structure of the xylooligosaccharide. In contrast, in the ricin 1α and 1β pockets (Figure 5a,c), the

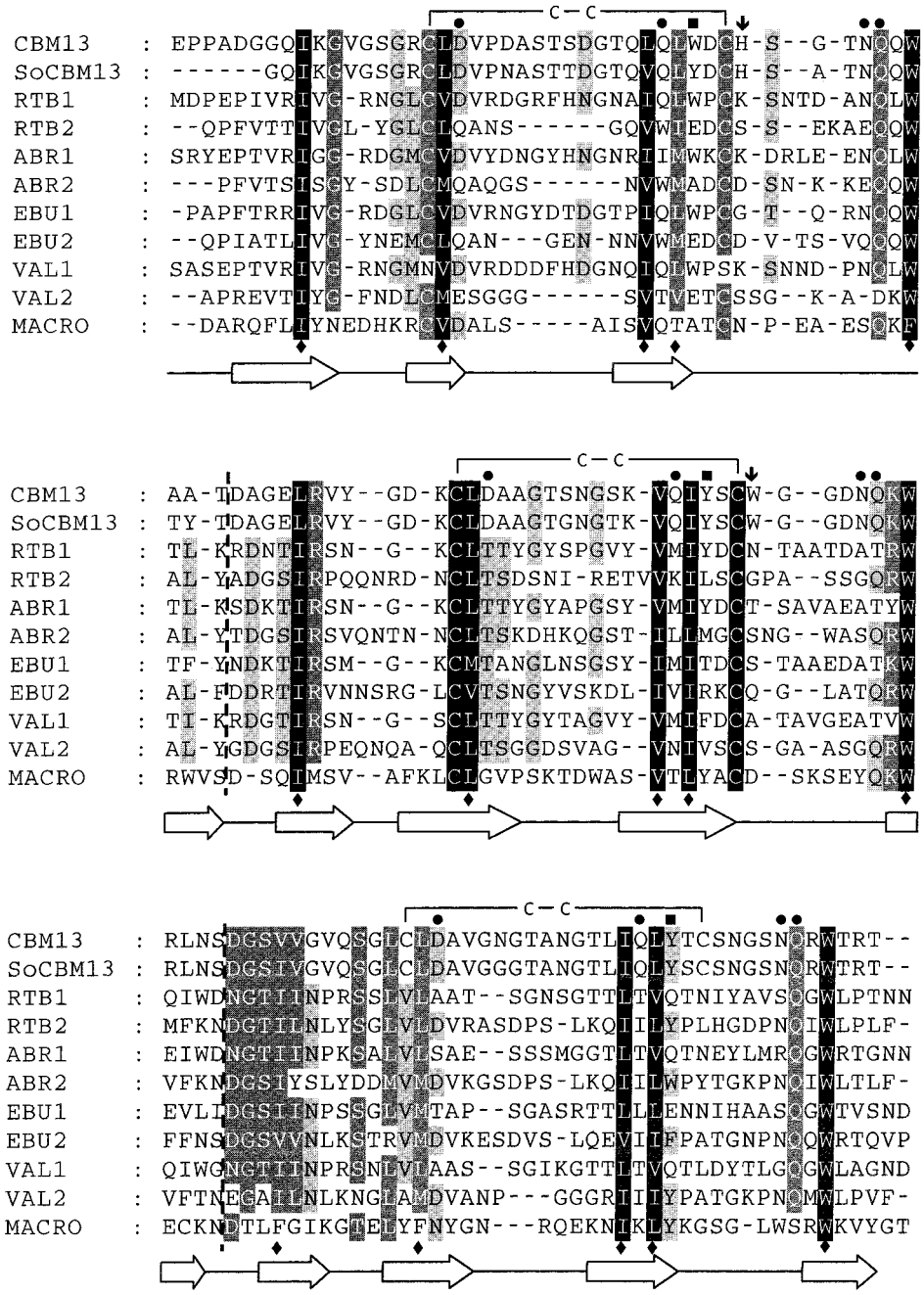

FIGURE 7: Amino acid sequence alignment based on the structures of β-trefoil modules known to bind carbohydrates: CBM13, family 13 CBM from *S. lividans* Xyn10A; SoCBM13, residues 313−436 of the *S. olivaceoviridis* Xyn10A structure (PDB entry 1XYF); RTB1 and RTB2, residues 5−136 and 138−262 of modules 1 and 2 of the *R. communis* toxin B chain, respectively (PDB entry 2AAI); ABR1 and ABR2, residues 10−141 and 144−267 of modules 1 and 2 of the *Abrus precatorius* agglutinin B chain, respectively (PDB entry 1ABR); EBU1 and EBU2, residues 8−136 and 138−264 of modules 1 and 2 of the *Sambucus ebulin* toxin B chain, respectively (PDB entry 1HWO); VAL1 and VAL2, residues 2−133 and 135−255 of modules 1 and 2 of *V. album* mistletoe lectin, respectively (PDB entry 2MML); and MACRO, residues 2−135 of the cysteine-rich macrophage mannose receptor (PDB entry 1FWV). Dashed vertical lines separate the α, β, and γ domains of the modules. Disulfide-bonded cysteine pairs in CBM13 are denoted above the sequences. Side chains in CBM13 hydrogen bonding with lactose are denoted with filled circles (●), and aromatic side chains interacting with lactose are denoted with filled squares (■). Additional residues in CBM13 found to interact with xylooligosaccharides are denoted with downward pointing arrows (↓). The secondary structure of CBM13, which very closely approximates that of all the shown sequences, is given below. Side chains forming the hydrophobic core in all of shown sequences are denoted below with filled diamonds (◆).

β-hairpin loop is extended and likely precludes longer polymers from binding in these sites due to steric hindrance and the absence of specific stacking interactions. On the other hand, the SoCBM13, which has also been reported to bind xylan (*26*), does contain these permutations in the α and β pockets, and presumably binds xylan polymers in a similar fashion. The absence of a ligand in the γ site of the CBM13−xylopentaose complex precludes comment on the architecture of this binding site. However, it is clear that it remains capable of binding xylooligosaccharides, as observed through NMR spectroscopy studies by Schärpf et al. (*5*), despite its lack of an additional planar amino acid side chain in the binding pocket.

Thus, the β-trefoil fold, and in particular the trefoil domain, appears to be an excellent general scaffold for carbohydrate recognition sites. To refer to this fold as a galactose binding

fold is misleading, as it is clearly capable of accommodating many sugars, making its evolutionary progenitor difficult to identify. While the majority of the currently characterized modules do recognize galactose, this is most likely a result of the early focus on examples from systems such as plants and animals, where the recognition of galactose on proteins and cell surfaces is important. As more microbial examples are characterized, it is likely that a more diverse binding specificity will continue to emerge and better represent the potential carbohydrate binding function of the β-trefoil fold. In light of this discussion, it seems doubtful that the β-trefoil folds of carbohydrate recognition proteins evolved from a galactose binding 42-amino acid polypeptide as proposed by Rutenber et al. (*3*). The ancestral polypeptide, should there have been one, was most likely of a more general binding specificity.

NOTE ADDED IN PROOF

Fujimoto et al. (*34*) have recently reported an X-ray crystallographic analysis of xylose, xylobiose, xylotriose, glucose, galactose, and lactose bound to the α- and γ-sites of the homologous family 13 xylan binding domain from *Streptomyces olivaceoviridis* E-86.

ACKNOWLEDGMENT

We thank N. Ghuman for optimizing complex crystallization conditions, Dr. D. A. Kuntz for help with crystallization and data collection, Dr. J. van den Elsen for molecular replacement trials, and A. Perrakis for help with ARP/wARP implementing Refmac5.

REFERENCES

1. Coutinho, P. M., and Henrissat, B. (1999) http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html.
2. 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.) The Royal Society of Chemistry, Cambridge, U.K.
3. Rutenber, E., Ready, M., and Robertus, J. D. (1987) *Nature 326*, 624−626.
4. Boraston, A. B., Tomme, P., Amandoron, E. A., and Kilburn, D. G. (2000) *Biochem. J. 350*, 933−941.
5. Schärpf, M., Connelly, G. P., Lee, G. M., Boraston, A. B., Warren, R. A. J., and McIntosh, L. P. (2002) *Biochemistry 41*, 4255−4263.
6. Ziser, L., Setyawati, I., and Withers, S. G. (1995) *Carbohydr. Res. 274*, 137−153.
7. Fischer, E., and Zemplen (1910) *Ber. Dtsch. Chem. Ges. 43*, 2536.
8. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W., Jr., and Sweet, R. M., Eds.) pp 307−326, Academic Press, New York.
9. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D55*, 849−861.
10. Cowtan, K. D., and Zhang, K. Y. (1999) *Prog. Biophys. Mol. Biol. 72*, 245−270.
11. Collaborative Computational Project, No. 4 (1994) *Acta Crystallogr. D50*, 760−763.
12. Jones, T. A., Bergdoll, M., and Kjeldgaard, M. (1990) in *Crystallographic and modeling methods in molecular design* (Bugg, C. E., and Ealick, S. E., Eds.) pp 63−72, Springer-Verlag, New York.
13. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
14. Kleywegt, G. J., and Brünger, A. T. (1996) *Structure 4*, 897−904.
15. Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
16. Perrakis, A., Morris, R. J., and Lamzin, V. S. (1999) *Nat. Struct. Biol. 6*, 458−463.
17. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D53*, 240−255.
18. Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J., and Williamson, M. P. (2000) *J. Biol. Chem. 275*, 41137−41142.
19. Bolam, D. N., Xie, H., White, P., Simpson, P. J., Hancock, S. M., Williamson, M. P., and Gilbert, H. J. (2001) *Biochemistry 40*, 2468−2477.
20. Notenboom, V., Boraston, A. B., Kilburn, D. G., and Rose, D. R. (2001) *Biochemistry 40*, 6248−6256.
21. Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A. J., and Kilburn, D. G. (2001) *Biochemistry 40*, 6240−6247.
22. Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J., and Fontes, C. M. (2000) *Biochemistry 39*, 5013−5021.
23. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., and Blow, D. M. (1974) *Biochemistry 13*, 4212−4228.
24. Murzin, A. G., Lesk, A. M., and Chothia, C. (1992) *J. Mol. Biol. 223*, 531−543.
25. Hazes, B. (1996) *Protein Sci. 5*, 1490−1501.
26. Kuno, A., Kaneko, S., Ohtsuki, H., Ito, S., Fujimoto, Z., Mizuno, H., Hasegawa, T., Taira, K., Kusakabe, I., and Hayashi, K. (2000) *FEBS Lett. 482*, 231−236.
27. Hirabayashi, J., Dutta, S. K., and Kasai, K. (1998) *J. Biol. Chem. 273*, 14450−14460.
28. Shimoi, H., Iimura, Y., Obata, T., and Tadenuma, M. (1992) *J. Biol. Chem. 267*, 25189−25195.
29. Shen, S. H., Chretien, P., Bastien, L., and Slilaty, S. N. (1991) *J. Biol. Chem. 266*, 1058−1063.
30. Leteux, C., Chai, W., Loveless, R. W., Yuen, C. T., Uhlin-Hansen, L., Combarnous, Y., Jankovic, M., Maric, S. C., Misulovin, Z., Nussenzweig, M. C., and Feizi, T. (2000) *J. Exp. Med. 191*, 1117−1126.
31. Liu, Y., Chirino, A. J., Misulovin, Z., Leteux, C., Feizi, T., Nussenzweig, M. C., and Bjorkman, P. J. (2000) *J. Exp. Med. 191*, 1105−1116.
32. Evans, S. V. (1993) *J. Mol. Graphics 11*, 134−138.
33. Evans, S. V. (1993) *J. Mol. Graphics 11*, 127−128.
34. Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2002) *J. Mol. Biol. 316*, 65−78.

BI015865J