Solution structure of family 21 carbohydrate-binding module from *Rhizopus oryzae* glucoamylase

Yu-Nan LIU*, Yen-Ting LAI*, Wei-I CHOU†, Margaret Dah-Tsyr CHANG†¹ and Ping-Chiang LYU*¹

*Institute of Bioinformatics and Structural Biology, National Tsing Hua University, No. 101, Sec. 2, Kuang Fu Rd, Hsinchu, Taiwan 30013, and †Institute of Molecular and Cellular Biology and Department of Life Science, National Tsing Hua University, No. 101, Sec. 2, Kuang Fu Rd., Hsinchu, Taiwan 30013

CBMs (carbohydrate-binding modules) function independently to assist carbohydrate-active enzymes. Family 21 CBMs contain approx. 100 amino acid residues, and some members have starch-binding functions or glycogen-binding activities. We report here the first structure of a family 21 CBM from the SBD (starch-binding domain) of *Rhizopus oryzae* glucoamylase (RoCBM21) determined by NMR spectroscopy. This CBM has a β -sandwich fold with an immunoglobulin-like structure. Ligand-binding properties of RoCBM21 were analysed by chemical-shift perturbations and automated docking. Structural comparisons with previously reported SBDs revealed two types of topologies, namely type I and type II, with CBM20, CBM25, CBM26 and CBM41 showing type I topology, with CBM21 and CBM34 showing type II topology. According to the chemical-shift perturbations,

RoCBM21 contains two ligand-binding sites. Residues in site II are similar to those found in the family 20 CBM from *Aspergillus niger* glucoamylase (AnCBM20). Site I, however, is embedded in a region with unique sequence motifs only found in some members of CBM21s. Additionally, docking of β -cyclodextrin and malto-oligosaccharides highlights that side chains of Y83 and W47 (one-letter amino acid code) form the central part of the conserved binding platform in the SBD. The structure of RoCBM21 provides the first direct evidence of the structural features and the basis for protein–carbohydrate recognition from an SBD of CBM21.

Key words: carbohydrate-active enzyme, carbohydrate-binding module (CBM), glucoamylase, *Rhizopus oryzae*, solution structure, starch-binding domain (SBD).

INTRODUCTION

Carbohydrates are the most diverse biopolymer in biological systems and manifest this diversity in several ways, including (1) the number of monomeric units in polymers, (2) the types of monosaccharides and (3) the kinds of glycosidic linkages. This diversity affords carbohydrates numerous critical biological roles [1] in energy and carbon storage, mechanical support and buffering of force or environmental stress [2,3], modulation of cell physiology, cellular signalling and molecular recognition [4]. For instance, cellulose is the major structural component of plant tissue, and chitin constitutes the exoskeleton of arthropods. Other examples include glucosaminoglycans, which comprise the extracellular ‘glue’ that cushions mechanical forces on joints [2], and xanthan gum, which contributes to the formation of bacterial biofilms that protect bacteria from environmental stress [3]. Starch and glycogen also function as energy sources for plants and animals respectively.

Starch is an abundant polysaccharide composed of two types of D-glucose polymers: amylose and amylopectin. The former is composed of mostly linear α -1,4-linked glucose residues, whereas the latter is the highly branched component of starch containing 5–6 % of α -1,6-linkages. Together amylose and amylopectin fold into helical structures [5,6] that are further organized into compact granular structures to efficiently save energy and the carbon for life. On the other hand, glycogen is a readily mobilized storage form of glucose in animals. The glucose residues in glycogen are

also linked by α -1,4-glycosidic bonds, and branched α -1,6-glycosidic bonds occur at about every tenth residue. The controlled breakdown of glycogen buffers the blood glucose level.

Biodegradation of polysaccharides allows the photosynthetically fixed energies and carbon to enter recycling pathway of ecosystem. Cells have developed various carbohydrate-active enzymes [7,8], such as glycoside hydrolases, glycosyltransferases, polysaccharide lyases and carbohydrate esterases, which modify polysaccharides to activate their biological functions. Several of these carbohydrate-active enzymes have developed into a modular structure comprising a catalytic module and one or more CBMs (carbohydrate-binding modules) and some ancillary modules. A CBM is defined as contiguous amino acid sequence (~50–200 amino acids) within a carbohydrate-active enzyme with a discrete fold having carbohydrate-binding activity. To date, 45 families of CBMs have been recorded on the Carbohydrate-Active Enzyme website (<http://afmb.cnrs-mrs.fr/CAZY/>).

CBM family 21 is a family of modules containing ~100 amino acid residues. This family currently has 63 entries: one bacterial, 29 fungal and 33 metazoal (<http://afmb.cnrs-mrs.fr/CAZY/>). The starch-binding function of the CBM from *Rhizopus oryzae* glucoamylase has been extensively studied [9,10]. Protein phosphatase-1s (PP1s), which respond to signals such as insulin [11], indirectly manipulate carbohydrates and control glycogen metabolism in eukaryotes. Moreover, regulation of this enzyme is strongly correlated with the cause of diabetes in humans [12]. The glycogen-binding functions of PP1G (PP1 regulatory subunit) in

Abbreviations used: AnCBM20, family 20 CBM from *Aspergillus niger* glucoamylase; BhCBM25 and BhCBM26, families 25 and 26 CBMs from *Bacillus halodurans* maltohexaose-forming amylase; BMRB, Biological Magnetic Resonance Data Bank; CBM, carbohydrate-binding module; 2D, two-dimensional; NOE, nuclear Overhauser effect; PDB, Protein Data Bank; PP1, protein phosphatase 1; PP1G, protein phosphatase-1 regulatory subunit; rmsd, root mean square deviation; RoCBM21, family 21 CBM from *Rhizopus oryzae* glucoamylase; SBD, starch-binding domain; TvCBM34 I and TvCBM34 II, family 34 CBMs from *Thermoactinomyces vulgaris* from α -amylase I and α -amylase II; for brevity the one-letter code is used for amino acid residues (e.g. Y83 is tyrosine-83).

¹ Correspondence may be addressed to either of these authors (email dtchang@life.nthu.edu.tw or pclyu@life.nthu.edu.tw).

The atomic co-ordinates and structure factors (2DJM) have been deposited in the PDB (Protein Data Bank), Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ, U.S.A. (<http://www.rcsb.org/>). The chemical shifts of RoCBM21 have been deposited in the BMRB (Biological Magnetic Resonance Data Bank), University of Wisconsin-Madison, Madison, WI, U.S.A.) under the entry number BMRB7083.

several higher eukaryotes also have been widely studied. PP1Gs bind glycogen via family 21 CBMs [13].

Seven of the 45 CBM families (20, 21, 25, 26, 34, 41 and 45) have starch-binding activity. Starch-binding CBMs are also called SBDs (starch-binding domains) in glycoside hydrolases. The dissociation constants (K_d) for the binding of SBDs to starch are in the micromolar range [9,14]. SBDs are functionally independent of the catalytic domains and are postulated to disrupt the structure of starch and to anchor the catalytic domain to insoluble starch [14–17]. Prior to our present study, no representative structure of a CBM21 has been available, and thus the detailed interactions between CBM21s and starch/glycogen remain unclear. Here we describe the first structure of a family 21 CBM from *R. oryzae* glucoamylase (*RoCBM21*) as determined by NMR spectroscopy. We compare this structure with that of several previously reported SBDs. Our results also illustrate the conservation of the starch-binding fold and the diverse mechanisms by which residues in SBDs interact with ligands.

EXPERIMENTAL

Clone construction

The gene encoding *RoCBM21* was cloned on the basis of a previously reported protocol with slight modification [9]. The DNA fragment encoding SBD (A1–T106; for brevity the one-letter code for amino acids is used) of *R. oryzae* glucoamylase was amplified by PCR using the forward primer 5'-CATATGGCAAGTATTCC-TAGCAGT-3' and the reverse primer 5'-CTCGAGTTATGTAG-ATACTTGGT-3' (restriction sites are in bold). The PCR product was cloned into the pGEM-T Easy cloning vector (Promega) and verified by DNA sequencing. The SBD DNA fragment was subsequently ligated into the pET23a(+) expression vector (Novagen) at NdeI and XhoI sites to generate pET-*RoCBM21*. There is a difference (T53 → I53) between the sequence in database (NCBI protein entry: ABB77799) and the *RoCBM21* that we report here because the gene for this *RoCBM21* was cloned from a local strain of *R. oryzae*.

Sample purification

Escherichia coli BL21-Gold (DE3) cells (Novagen) transformed with pET-*RoCBM21* were grown in Luria–Bertani medium containing ampicillin (100 µg/ml) at 37 °C with shaking at 250 rev./min until the attenuation (D_{600}) reached 0.6. For isotope labelling, BL21-Gold (DE3) cells with pET-*RoCBM21* were cultured in M9 medium [18] with $^{15}\text{NH}_4\text{Cl}$ as nitrogen source and ^{13}C glucose as carbon source. Protein synthesis was induced by adding isopropyl β -D-thiogalactoside to a final concentration of 400 µM, and the cell culture was incubated at 20 °C for 16 h. The cells were harvested by centrifugation at 3700 g for 15 min at 4 °C, and the pellet was resuspended in 20 ml of 10 mM sodium acetate, pH 4.5, and then sonicated. The cell debris was removed by centrifugation at 16000 g for 15 min at 4 °C. The supernatant was applied to amylose resin (New England BioLabs) (equilibrated and washed with 10 mM sodium acetate, pH 4.5) and was eluted with 10 mM glycine, pH 10. The eluate was dialysed against 10 mM sodium acetate, pH 4.5, using an Amicon concentrator (M_r cut-off 3000). The concentrated sample was further subjected to a Hitrap SP column (GE Healthcare) and washed with sodium acetate buffer, pH 4.5, containing 50 mM NaCl.

NMR spectroscopy for structure determination

NMR data were acquired on a Bruker Avance 600 MHz or 800 MHz spectrometer. For structure determination, 1 mM *RoCBM21* (unlabelled, ^{15}N -labelled or $^{13}\text{C}/^{15}\text{N}$ -double-la-

belled) was dissolved in 10 mM sodium acetate, pH 4.5, and subjected to NMR experiments at 25 °C. The protein concentrations were quantitated by Bio-Rad Protein Assay. Backbone assignment was accomplished with HNCA [19–21], HN(CO)CA [19,21], HNCACB [22], CBCA(CO)NH [23,24], HNCO [19–21] and HN(CA)CO [21] experiments [25]. Because *RoCBM21* contains a relatively high proportion of aromatic residues, the assignment of aromatic side chains was assisted by HBCBCGCDHD and HBCBCGCDCEHE experiments [26]. The chemical-shift resonance assignments of remaining atoms were accomplished using both NOESY [27,28] and ^1H - ^{15}N HSQC (heteronuclear single-quantum coherence)-NOESY [29,30] with the assistance of through-bond correlation spectra. Homonuclear DQF (double quantum-filtered)-COSY [31], TOCSY [27] and ^1H - ^{15}N HSQC-TOCSY were utilized to obtain through-bond correlations. The mixing times were as follows: TOCSY spectra, 90 ms; NOESY spectra, 50, 100 or 150 ms. All two-dimensional (2D) spectra were recorded with 512 t1 increments, and 2048 t2 complex data points were processed using TopSpin 1.3 (Bruker). Distance restraints were derived from NOESY spectra recorded with a 100 ms mixing time. A 2D ^1H - ^{15}N HSQC spectrum was recorded after dissolving the freeze-dried *RoCBM21* in 99 % $^2\text{H}_2\text{O}$ at 25 °C for 36 h to identify the protected amide protons. *RoCBM21* did not readily dissolve after freeze-drying, and therefore excess $^2\text{H}_2\text{O}$ was added to completely dissolve the protein. The excess $^2\text{H}_2\text{O}$ was subsequently removed by freeze-drying to yield a 500 µl sample. Hydrogen-bond constraints were obtained from the HSQC amide proton protection and confirmed with surrounding NOE (nuclear Overhauser effect) signals and HNCOHB (through-hydrogen-bond coherence) [32]. Dihedral angle constraints were obtained using the TALOS program [33] with chemical shifts of N, HA, CA, CB and C atoms. Sodium 2,2-dimethyl-2-silapentane-5-sulphonate was used as an internal reference for proton chemical shifts, and heteronuclear chemical shifts were referenced assuming $\gamma^{15}\text{N}/\gamma^1\text{H} = 0.101329118$ and $\gamma^{13}\text{C}/\gamma^1\text{H} = 0.251449530$.

Structure calculation and structural analyses

Partially assigned peak lists and chemical-shift lists were acquired from manual assignment using the program SPARKY 3 (<http://www.cgl.ucsf.edu/home/sparky/>). The peak intensities were derived using the default peak-fitting protocol assuming Lorentzian lineshapes. Structure calculations for *RoCBM21* were carried out using CNS 1.1 [35] and ARIA 2.0 [36] with torsion angle dynamics and standard simulated annealing protocols [37]. These calculations were followed by explicit water refinement using the OPLS force field [38]. Of 200 structures that we obtained, the 15 structures with the lowest total energies were selected for analysis. Their quality was assessed with PROCHECK-nmr [39]. The SBD structures for comparison are CBMs from *A. niger* glucoamylase (*AnCBM20*) [40], *Thermoactinomyces vulgaris* R-47 α -amylase I (*TvCBM34 I*) [41], *Bacillus halodurans* maltohexaose-forming amylase (*BhCBM25* and *BhCBM26*) [42] and *Klebsiella pneumoniae* pullulanase (*KpCBM41*) [43].

Chemical-shift perturbations

To investigate the ligand-binding residues and ligand-binding interactions of *RoCBM21*, maltotriose, maltoheptaose and β -cyclodextrin were applied to the chemical-shift perturbation. Maltotriose and maltoheptaose were prepared as 100 mM stocks, whereas β -cyclodextrin was prepared as a 20 mM stock, because of its lower solubility. *RoCBM21* (1 mM) was titrated with individual ligands, and the ^1H - ^{15}N HSQC spectra of *RoCBM21* were

Figure 1 NMR spectra of RoCBM21

(A) Assignment of amide resonances of RoCBM21 on a ¹H-¹⁵N HSQC spectrum. All backbone amide peaks are well resolved, except for Q10, which overlaps D12, and Y86, which overlaps E87. (B) Example assignment strips from residue F21 to Y26 in the HNCACB spectrum. C α peaks are phased to the positive phase (black); C β , negatively phased peaks (grey). (C, D) Antiparallel secondary structures with a bulged structure in the N-sheet and a loop in C-sheet. Thick arrows indicate the NOEs found between the sequential H α and the amide proton, thick double-headed arrows indicate interstrand H α -H α NOEs, thin double-headed arrows indicate interstrand amide proton to amide proton NOEs, thin arrows indicate interstrand H α to amide proton NOEs and dotted lines indicate interstrand hydrogen bonds. NOEs in the loop regions are not shown.

recorded to monitor the interactions. Weighted averaged ¹H and ¹⁵N chemical-shift changes were calculated using the following equation [44–47]:

$$\Delta\delta_{\text{avg}} = \{[(\Delta\delta_{\text{H}})^2 + (0.2\Delta\delta_{\text{N}})^2]/2\}^{1/2} \quad (1)$$

Docking simulations

AutoDock3.05 [48] was used to simulate the binding modes of starch molecules at the binding sites. The carbohydrate molecules

were docked to different binding sites in separate simulations. Affinity grids, 90 × 90 × 90, three-dimensionally centred on the binding sites with 0.375 Å (1 Å = 0.1 nm) spacing were calculated using the program Autogrid [48]. The Lamarckian genetic algorithm was used for conformational searches. For each carbohydrate at one of the two binding sites, 100 trials were made with a population size of 150 for each trial. Initial position and conformation were chosen randomly. The translation step was 2.0 Å, and the rotation step was 50°. Other docking parameters were as follows: mutation rate, 0.02; cross-over rate, 0.8; elitism, 1;

Figure 2 Solution structure of *RoCBM21*

(A) Stereo view of the *RoCBM21* ensemble. The front side of *RoCBM21* is shown with the N-terminal loop up and C-terminal loop down. Strands are in cyan. (B, C) Ribbon view and surface view of *RoCBM21* respectively.

local search rate, 0.06; with 1 million energy evaluations. Final conformations from the 100 trials were clustered using an rmsd (root mean square deviation) tolerance of 1.5 Å.

RESULTS

NMR spectra and molecular structure

The ^1H – ^{15}N HSQC spectrum of *RoCBM21* showed a well-dispersed pattern of peaks characteristic of typical β -strands (Figure 1A). The spectral widths of proton dimensions were set to 16 p.p.m. to accommodate the peaks. The most upfield chemical shift was -1.181 p.p.m., corresponding to I79 H δ 1; the H ϵ 1 of W70 resulted in the most downfield chemical shift at 11.88 p.p.m. Owing to the large number of aromatic residues in *RoCBM21* (18/106), several chemical shifts were affected by the π -electron currents of aromatic rings. An atom that is close to an aromatic ring may experience either shielding (above or below the ring) or deshielding (in the plane of the aromatic ring) effect of the π -electron current. Some assigned chemical shifts have even been reported as anomalous or suspicious by the software currently used by BMRB to check for chemical-shift outliers [49]. These chemical shifts were carefully verified on request of the BMRB annotator. For example, the chemical shift of V39 H β (0.739) may be affected by the rings of Y83 and W47, N52 H β 3 (0.775) by the ring of W47, N97 H β 3 (-0.09) by the ring of Y102, I79 H γ 12 (-0.22) by the ring of Y40, and Y102 H β 3 (0.482) by

the ring of F82; as a result, their corresponding resonances have chemical-shift values upfield-shifted from those normal expected for the listed atoms. In HNCA, HN(CO)CA, CBCA(CO)NH, HNCACB, HNCO, and HN(CA)CO spectra, the backbone sequential connectivity proceeds continuously through the whole protein sequence, except for proline residues and G18. The connectivity between D17, G18 and S19 was absent in all those backbone connective experiments, and thus the chemical shift of amide nitrogen from G18 could not be assigned unambiguously. An example view of strips demonstrating backbone connectivity in an HNCACB spectrum is shown in Figure 1(B).

The structures were calculated based on 2247 restraints, including 2071 NOE-derived distance constraints, 102 dihedral angle restraints, and 74 distance restraints from hydrogen bonds as described in the Experimental section. NOEs of N- and C- β -sheets are illustrated in Figure 1(C). A total of 200 structures were generated at the final iteration of the ARIA software calculation. A total of 15 structures with the lowest total energies were chosen for analyses and were deposited in the PDB (Figure 2A). The structure with the lowest rmsd to the average structure was chosen as the representative structure (Figure 2B). The NMR statistics are summarized in Table 1. The RMSD with respect to the average structure was 0.48 ± 0.06 Å for backbone and 0.96 ± 0.11 Å for heavy atoms in the well-defined region (residues 8–16, 20–40, 52–60, 67–87, 92–95 and 102–104) and 1.14 ± 0.31 Å for backbone and 1.43 ± 0.29 Å for heavy atoms for all residues. In the Ramachandran plot, 95 % of non-glycine and non-proline

Table 1 Structural statistics of *RoCBM21* in aqueous solution at pH 4.5 and 298 K

The well-defined region refers to regions including residues 8–16, 20–40, 52–60, 67–87, 92–95 and 102–104. Note: 1 kcal = 4.184 kJ.

Parameter	Value
Experimental constraints	
Total	2071
Intra-residue	973
Sequential ($ i-j = 1$)	509
Medium-range ($ i-j \leq 4$)	142
Long-range ($ i-j > 4$)	447
Hydrogen-bond constraints	74
Dihedral-angle constraints	102
rmsd from experimental data	
Distance (Å)	0.0557 ± 0.0077
rmsd (Å) with respect to the average structure	
Well-defined region	
Backbone (N, Cα and C')	0.48 ± 0.06
Heavy atoms	0.96 ± 0.11
All residues	
Backbone	1.14 ± 0.31
Heavy atoms	1.43 ± 0.29
Total energy E_{total} after water refinement (kcal · mol ⁻¹)	-2730 ± 66
rmsd from idealized covalent geometry	
Bonds (Å)	0.0046 ± 0.0004
Angles (°)	0.546 ± 0.035
Impropers (°)	0.522 ± 0.046
Ramachandran analysis	
Residues in most favoured regions	67.3%
Residues in additionally allowed regions	27.7%
Residues in generously allowed regions	3.5%
Residues in disallowed regions	1.5%

residues are in the most-favoured or additionally allowed region, and 98.5 % are in the generously allowed region. Most of the N97 residues and several of the N45 and N101 residues of the ensemble were found in the disallowed region. N97 and N101 are located in loop 8, whereas N45 is in loop 4; both loops are in the most flexible region of *RoCBM21* (see Supplementary Figure 1 at <http://www.BiochemJ.org/bj/403/bj4030021add.htm>).

The *RoCBM21* domain contains 106 residues, the sequence of which has little similarity (< 25 %) to that of other SBD families. The solution structure of *RoCBM21*, however, shows a conventional β -sandwich fold and an immunoglobulin-like architecture, features that are characteristic of most CBMs [17]. The β -sandwich is symmetric and is composed of eight antiparallel β -strands: $\beta 1$ (V9–Y16), $\beta 2$ (F21–V27), $\beta 3$ (V34–D42), $\beta 4$ (I53–

G60), $\beta 5$ (Y67–A74), $\beta 6$ (I79–V88), $\beta 7$ (T92–N95), and $\beta 8$ (Y102–V104). These β -strands can be subdivided into an N-terminal strand containing β -sheet (N-sheet) (Figure 2B), consisting of $\beta 1\beta 2\beta 5$, and a C-terminal strand containing β -sheet (C-sheet) (Figure 2C), consisting of $\beta 3\beta 6\beta 7/8$, which are connected by $\beta 4$. Strand $\beta 2$, positioned in the middle of the N-sheet, is paired antiparallel to $\beta 1$ and $\beta 5$, whereas $\beta 6$, positioned in the middle of the C-sheet, is paired antiparallel to $\beta 3$, $\beta 7$ and $\beta 8$. Strand $\beta 4$, which is partially paired antiparallel to $\beta 3$ and $\beta 5$, lies across both β -sheets. The hydrophobic core of *RoCBM21* is composed of V9, L11, I14, Y16, F21, I25, V27, W70 and F72 in the N-sheet and V36, V38, Y40, F82, I84, Y86, V88, Y93, Y102 and V104 in the C-sheet. In $\beta 1$, residues L11–Y14 are not hydrogen-bonded to $\beta 2$, and they form a bulge (Figure 1C). Both $\beta 7$ and $\beta 8$ are hydrogen-bonded to $\beta 6$ and they are spanned by loop 8 (Figure 1D). The solvent-accessible surface of the *RoCBM21* structure is shown in Figure 2(C). The DelPhi electrostatic potential [50] was mapped on to the surface.

Structural comparisons with SBDs

We compared the *RoCBM21* structure with those of different SBD families (Table 2). Figures 3(A), 3(B) and 3(C) show the similarity in primary, secondary and tertiary structures respectively. Equivalent β -strands and loops in different families of SBDs are labelled and grey-scale-coloured according to the structure of *RoCBM21* (Figure 3A). Two types of topologies can be discerned from the structural comparison: the structures of *AnCBM20*, *BhCBM25*, *BhCBM26* and *KpCBM41* have type I topology, whereas the structures of *RoCBM21s* and *TvCBM34* structures show type II topology. These two topology types are similar, except that a strand must be shifted to overlap two topologies. For example, $\beta 1$ in *AnCBM20* is equivalent to $\beta 2$ in *RoCBM21*; subsequent strand equivalents can be fitted, one by one, and the final strand ($\beta 8$) of *RoCBM21* is superimposable with strand 7 of *AnCBM20*. Notably, the final strand in *AnCBM20* plays the same role as $\beta 1$ in *RoCBM21*, forming hydrogen bonds with the middle strand ($\beta 2$ or its equivalent) of the N-terminal β -sheet. All β -strands in *RoCBM21* are antiparallel, but the first and the last strands are parallel in *AnCBM20*. In sum, the overall topology of *RoCBM21* (type II topology) is similar to that of *AnCBM20* (type I topology), but the order of the equivalent strands is sequentially shifted by one. It appears that most N-terminal SBDs have type II topology, whereas the C-terminal SBDs have type I topology (except for *KpCBM41*) (Table 2) [43].

Ligand-binding and chemical-shift perturbation

Three carbohydrates (maltotriose, maltoheptaose and β -cyclodextrin) tested in the present study showed similar patterns of chemical-shift perturbation; the same amino acid residues were affected

Table 2 Structural comparison of different families of SBDs with *RoCBM21*

The sequence ranges in PDB files of *BhCBM25* and *BhCBM26* correspond to amino acids 863–958 and 771–863 of open-reading-frame BH0413.

Parameter	Family ...	Value					
		<i>RoCBM21</i>	<i>TvCBM34</i>	<i>AnCBM20</i>	<i>BhCBM25</i>	<i>BhCBM26</i>	<i>KpCBM41</i>
Module arrangement (N- or C-terminal)		N	N	C	C	C	N
PDB code		2DJM	1UH3	1KUL	2C3X	2C3G	2FHF
Sequence range (by PDB numbering)		1–106	1–123	509–616	10–102	1–98	53–155
Sequence identity (%)		100	10.6	7.9	13.2	8.3	6.3
Sequence similarity (%)		100	21.6	16.7	20.8	20.0	14.3
Topology type		II	II	I	I	I	I

Figure 3 Type I topology and type II topology of SBDs

The regions of the eight β -strands in RoCBM21 and their corresponding strands in other SBDs are shown. Strands 7 and 8 form hydrogen bonds with strand 6. The extra loop between the first two β -strands (corresponding to the bulged structure of RoCBM21) at the N-terminus of TvCBM34 I is underlined. (A) Primary structures of SBDs. The sequences corresponding to secondary structure are indicated. (B) Schematic diagrams of type I topology and type II topology. The strand order is equivalent to that shown in (A) and is shown above the type-II-topology diagram. (C) Three-dimensional structures of type I topology (represented by AnCBM20) and type II topology (represented by RoCBM21).

in the titrations, but the magnitude of the change differed between carbohydrates. The ^1H – ^{15}N HSQC spectra of RoCBM21 before and after titration with β -cyclodextrin overlapped (Figure 4A), and a summary of the chemical-shift changes and the residues affected is plotted in Figure 4(B). The RoCBM21 residues that exhibited significant chemical-shift perturbations (1 S.D.: 0.07 p.p.m. as a threshold for β -cyclodextrin and maltoheptaose, and 0.03 p.p.m. for maltotriose) were mapped on the three-dimensional structure (Figure 4C). According to these perturbations, the residues affected by ligand binding can be categorized into three types. First, residues A41, W47, N52, Y83, K85, K91, D95, N96, N97 and S99 are located at the corresponding site I of previously reported SBDs. Likewise, residues N29, I30, A31, Y32, S33, K34, S57, F58, I62, N66, Y67, E68 and Y69 form the corresponding site II. The residues with significant chemical-shift changes and two carbohydrate-binding sites were mapped on the structure of RoCBM21 (Figure 4C). Interestingly, several residues located in the hydrophobic core were also affected by ligand binding; they are L11, V36, V38, W70 and I79. Loops 1, 4 and 8 are flexible regions with average rmsd values > 1.5 Å (see Supplementary Figure 1 at <http://www.BiochemJ.org/bj/403/bj4030021add.htm>). Besides high flexibility, loops 4 and 8, which enclose site I, share another feature: they are rich in

asparagine residues (N46, N48, N49, N50 and N52 are in loop 4, and N96, N97, N98 and N101 in loop 8). The titration with carbohydrate ligands caused large chemical-shift perturbations in asparagine residues N50, N52, N96, N97 and N98, suggesting that residues in these polyasparagine loops might be involved in CBM–starch interactions. However, the specific role of these polyasparagine loops still needs further studies. The presence of these polyasparagine loops is a distinct feature of some members of CBM21 [51]. Two molecules of β -cyclodextrin are docked respectively into the site I and site II of RoCBM21 (Figure 4D). In site I, the side chains of W47 and Y83 form the central part of the binding platform [42] (play roles as H26, W34 and W74 in CBM25, and Y23, Y25 and W36 in CBM26). These two aromatic side chains undergo hydrophobic interaction (we avoid the term ‘stacking’ for fear of confusing organic chemists) with the ring of glucose residues and collaborate with residues surrounding site I to bind ligands. Docking of maltotriose and maltoheptaose (see Supplementary Figure 3 at <http://www.BiochemJ.org/bj/403/bj4030021add.htm>) reveals that maltoheptaose interacts with W47 and Y83 in a similar pattern to β -cyclodextrin; maltotriose, on the other hand, is further away from W47. This consists with the smaller chemical-shift perturbations to loop-4 residues in maltotriose titration. In site II, residues N29, Y32 and E68 (which

Figure 4 Ligand-binding and ligand-docking studies of *RoCBM21*

(A) The ^1H – ^{15}N HSQC spectrum of *RoCBM21* before (black peaks) and after (red peaks) titration (β -cyclodextrin as ligand). Peaks with larger chemical-shift perturbations are indicated with green arrows. (B) Weighted averaged chemical-shift changes with respect to residue number. Black, light yellow and red represent the ligands maltotriose, maltoheptaose and β -cyclodextrin respectively. The perturbation thresholds are set to > 0.07 for β -cyclodextrin and maltoheptaose and 0.03 for maltotriose (dotted lines). (C) *RoCBM21* structure labelled with residues that are affected upon titration. Residues with chemical-shift changes larger than thresholds are considered as significantly affected (and thus hypothesized to play critical roles in ligand binding) are shown as stick structures. (D) Two molecules of β -cyclodextrin are docked into *RoCBM21*.

collaborate with residues surrounding site II) interact with three ligands similarly, as illustrated by the large chemical-shift changes observed. The dissociation constants (K_d) were extracted from fitting chemical-shift perturbations based on a theoretical curve [14,52,53]. The K_d fitted with site-I residues N50, W47

(side chain) and Y83 chemical-shift perturbations was determined to be $23.5 \pm 2.5 \mu\text{M}$; this value is close to that reported for site I of CBM20 [14,53], but is slightly higher than the value reported by Chou et al. [9]. The K_d fitted with site-II residues N29, I30 (Y32 was not used because of peaks missing during the

low-concentration titration) and E68 chemical-shift perturbations was calculated as $12.5 \pm 1.3 \mu\text{M}$, in agreement with the previously reported value [9].

DISCUSSION

We report here the first solution structure of a family-21 CBM. On the basis of the structures currently available, topologies of the SBDs can be categorized into two types. The type II topology contains all antiparallel β -strands, as exemplified by CBM21 and CBM34. Their first strands are hydrogen-bonded to their second strands, which are the centre strands of the N-sheets. In the type I topology, the first strands are the centre strands of the N-sheets; the tail strands take the place of the first strands in type II topology group.

Aromatic residues and polar residues play a vital functional role in CBMs [54–57]. Aromatic residues are thought to undergo hydrophobic interaction with the pyran rings of carbohydrates. The roles of aromatic residues in *RoCBM21*–carbohydrate binding have been studied by site-directed mutagenesis [9]. Mutants Y32A, Y67A, Y83A, W47A or Y93A significantly alter the starch absorption. Chemical-shift perturbations of Y32, Y67, Y83 and W47 were also observed in ^1H – ^{15}N HSQC, consistent with results obtained in mutagenesis studies. There was no chemical-shift perturbation for Y93, but its side chain was oriented toward the core of *RoCBM21*. Therefore mutation at this residue might affect *RoCBM21* stability and thus indirectly affect carbohydrate binding. Y32, located in loop 3 of site II, had the largest chemical-shift change. Interestingly, Y67 is close to site II, but its side chain points to the outside of the N-sheet rather than to site II, and there are corresponding outwardly oriented aromatic residues in several other SBDs (e.g. CBM34, CBM25, and CBM26). The significance of these conserved outwardly oriented aromatic residues remains unclear. Both Y83 and W47 are in site I, with Y83 in the middle of the C-sheet and W47 in loop 4. The observation of chemical-shift changes of both residues consists with the effect of mutations. This has revealed the involvement of these two residues in ligand binding. The side chain of W47 shows a chemical-shift change upon soluble ligand binding (Supplementary Figure 2 at <http://www.BiochemJ.org/bj/403/bj4030021add.htm>). This highlights the importance of such a tryptophan ring in starch binding.

C-terminal deletions have been made for *A. awamori* glucoamylase (97% similar to *A. niger* glucoamylase) [58]. Deletion of residues 597–615 does not affect the raw-starch absorption. This removed fragment 597–615 corresponds to *RoCBM21* β 1 and part of β 8 (Figures 3B and 3C). The rest of the truncated SBD is equivalent to β 2 in loop 8, and this region is invariant in both the type I and type II topologies of SBDs, implying a minimal scaffold for raw-starch absorption. This scaffold is also found in the N3 domain of *K. pneumoniae* pullulanase, although no starch-binding activity was reported. The N3 domain is a CBM-like domain that folds into a type-II-like topology. Further removal of residues 571–596 abolishes the raw-starch absorption, but retains some ability to assist the catalytic domain in the digestion of raw starch. The remaining part of the SBD is equivalent to the structure of the fragment from β 2 to β 5 of SBDs (Figures 3B and 3C). This result defines the minimal scaffold with which the binding domain can contribute to raw-starch digestion. The N-domain of isoamylase (PDB code: 2BF2) [59], which is not classified as a CBM, also contains such a functional scaffold. Deletion of residues 559–570 is equivalent to the removal of β 5. This deletion includes the critical core tryptophan residue, W562 (W70 in *RoCBM21*), and the disruption of the hydrophobic core results in a loss of the minimally required structural scaffold.

Our molecular structure of *RoCBM21* increases the understanding of the conserved folds of SBDs and the differences between these domains. Several mammalian proteins contain glycogen-binding family 21 CBMs, including the PP1G (human), suggesting a critical role for this family in glycogen synthesis and sugar metabolism [11]. The *RoCBM21* structure provides the basis for modelling other glycogen-binding family-21 CBMs. A better model will help to define a detailed molecular mechanism by which PP1G regulates glycogen synthesis and carbohydrate metabolism. However, a molecular model of PP1G built by Souchet et al. [60] (based on homology modeling of the ~ 300 -residue protein fragment) deviates substantially from our *RoCBM21* structure in the theoretical CBM21 region. Whether the CBM21 in PP1G compromises the overall similarity to the ~ 300 -residue protein fragment (maintaining a modular structure, as in *RoCBM21*) or compromises the modular homology {forming a part of the $(\alpha/\beta)_8$ -barrel, as Souchet et al. [60] proposed} is an interesting question worthy of further investigation.

Carbohydrates are important materials for the food and beverage industries. Perhaps, in the near future, they will also play significant roles in the fuel industry. Detailed knowledge of both carbohydrate manipulation, via studies of catalytic domains of carbohydrate-active enzymes, and carbohydrate binding, via studies of CBMs, is vital to an overall understanding of carbohydrate-active enzymes. The structural investigations of various families of SBDs and different members of CBM21s have allowed us to characterize inter-/intra-family variations of CBMs and different means by which these proteins recognize carbohydrates. These studies therefore may give the ability to engineer novel carbohydrate-active enzymes to meet future needs of humans.

The 800 MHz NMR spectra were obtained at High-Field Biomacromolecular NMR Core Facility (Taipei City, Taiwan) supported by the National Research Program for Genomic Medicine. The *RoCBM21* clone was kindly given by Mr Chia-Chin Sheu at Simpson Biotech Co. (Taipei, Taiwan). This work was supported by the National Science Council, Taiwan, R. O. C. (grant numbers 95-2752-B-007-003-PEA and 94-2311-B-007-005) and The National Tsing Hua University Structural Proteomic Center for Drug Discovery (Hsinchu, Taiwan) (grant 95N2517E1 to P.-C. L. and 95-2627-B-007-003 to M. D.-T. C.). We thank Dr Shou-Lin Chang (Department of Life Sciences, National Tsing Hua University, Hsinchu, Taiwan), Dr Shang-Te Danny Hsu (Bijvoet Center, Utrecht University, Utrecht, The Netherlands) and Dr Chi-Fon Chang (NMR Supporting Team, High-Field Biomacromolecular Solution NMR core facility, Institute of Biomedical Sciences, Taipei, Taiwan) for discussing NMR experiments, Ms Tsun-Ai Yo and Ms Nan-Lu Ho (National Tsing Hua University, Hsinchu, Taiwan) for technical help, and Dr Bunzo Mikami (Division of Applied Life Sciences, Graduate School of Agriculture, Kyoto University, Gokasho, Uji, Kyoto, Japan) for sharing files and information about CBM41.

REFERENCES

- Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G. and Marth, J. (1999) *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Kelly, G. S. (1998) The role of glucosamine sulfate and chondroitin sulfates in the treatment of degenerative joint disease. *Altern. Med. Rev.* **3**, 27–39.
- Crossman, L. and Dow, J. M. (2004) Biofilm formation and dispersal in *Xanthomonas campestris*. *Microbes Infect.* **6**, 623–629.
- Nagai, Y. (1998) Cell regulatory function of glycosphingolipids: carbohydrate recognition and biosignaling. *Pure Appl. Chem.* **70**, 49–53.
- Gessler, K., Usón, I., Takaha, T., Krauss, N., Smith, S. M., Okada, S., Sheldrick, G. M. and Saenger, W. (1999) V-amylase at atomic resolution: X-ray structure of a cycloamylose with 26 glucose residues (cyclomaltotetrahexaicosose). *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4246–4251.
- Smith, A. M., Denyer, K. and Martin, C. R. (1995) What controls the amount and structure of starch in storage organs? *Plant Physiol.* **107**, 673–677.
- Henrissat, B., Coutinho, P. M. and Davies, G. J. (2001) A census of carbohydrate-active enzymes in the genome of *Arabidopsis thaliana*. *Plant Mol. Biol.* **47**, 55–72.

- 8 Davies, G. J. and Henrissat, B. (2002) Structural enzymology of carbohydrate-active enzymes: implications for the post-genomic era. *Biochem. Soc. Trans.* **30**, 291–297
- 9 Chou, W. I., Pai, T. W., Liu, S. H., Hsiung, B. K. and Chang, M. D. (2006) The family 21 carbohydrate binding module of glucoamylase from *Rhizopus oryzae* consists of two sites playing distinct roles in ligand binding. *Biochem. J.* **396**, 469–477
- 10 Takahashi, T., Kato, K., Ikegami, Y. and Irie, M. (1985) Different behavior towards raw starch of three forms of glucoamylase from a *Rhizopus* sp. *J. Biochem. (Tokyo)* **3**, 663–671
- 11 Dent, P., Lavoie, A., Nakiely, S., Caudwell, F. B., Watt, P. and Cohen, P. (1990) The molecular mechanism by which insulin stimulates glycogen synthesis in mammalian skeletal muscle. *Nature* **348**, 302–308
- 12 Xia, J., Scherer, S. W., Cohen, P. T., Majer, M., Xi, T., Norman, R. A., Knowler, W. C., Bogardus, C. and Prochazka, M. (1998) A common variant in PPP1R3 associated with insulin resistance and type 2 diabetes. *Diabetes* **47**, 1519–1524
- 13 Armstrong, C. G., Doherty, M. J. and Cohen, P. T. (1998) Identification of the separate domains in the hepatic glycogen-targeting subunit of protein phosphatase 1 that interact with phosphorylase α , glycogen and protein phosphatase 1. *Biochem. J.* **336**, 699–704
- 14 Giardina, T., Gunning, A. P., Juge, N., Faulds, C. B., Furniss, C. S., Svensson, B., Morris, V. J. and Williamson, G. (2001) Both binding sites of the starch-binding domain of *Aspergillus niger* glucoamylase are essential for inducing a conformational change in amylose. *J. Mol. Biol.* **313**, 1149–1159
- 15 Southall, S. M., Simpson, P. J., Gilbert, H. J., Williamson, G. and Williamson, M. P. (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. *FEBS Lett.* **447**, 58–60
- 16 Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P. and Gilbert, H. J. (1998) *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem. J.* **331**, 775–781
- 17 Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781
- 18 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1995) Short Protocols In Molecular Biology, Wiley, New York
- 19 Grzesiek, S. and Bax, A. (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**, 432–440
- 20 Schleucher, J., Sattler, M. and Griesinger, C. (1993) Coherence selection by gradients without signal attenuation: application to the three-dimensional HNCQ experiment. *Angew. Chem. Int. Ed. Engl.* **32**, 1489–1491
- 21 Kay, L. E., Xu, G. Y. and Yamazaki, T. (1994) Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. *J. Magn. Reson. Ser. A* **109**, 129–133
- 22 Wittekind, M. and Mueller, L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the α - and β -carbon resonances in proteins. *J. Magn. Reson. Ser. B* **101**, 201–205
- 23 Grzesiek, S. and Bax, A. (1993) Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$ -enriched proteins. *J. Biomol. NMR* **3**, 185–204
- 24 Muhandiram, D. R. and Kay, L. E. (1994) Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. Ser. B* **103**, 203–216
- 25 Cavanagh, J., Fairbrother, W. J., Palmer, G. A. and Skelton, N. J. (1996) Protein NMR Spectroscopy, Academic Press Inc., New York
- 26 Yamazaki, T., Forman-Kay, J. D. and Kay, L. E. (1993) Two-dimensional NMR experiments for correlating $^{13}C\beta$ and $^1H\delta/\epsilon$ chemical shifts of aromatic residues in ^{13}C -labeled proteins via scalar couplings. *J. Am. Chem. Soc.* **115**, 11054–11055
- 27 Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2**, 661–665
- 28 Sklenar, V., Piotto, M. and Saudek, V. (1993) Gradient-tailor water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson. Ser. A* **102**, 241–245
- 29 Palmer, G. A., Cavanagh, J., Wright, P. E. and Rance, M. (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson.* **93**, 151–170
- 30 Kay, L. E., Keifer, P. and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665
- 31 Derome, E. A. and Williamson, M. P. (1990) Rapid-pulsing artifacts in double-quantum-filtered COSY. *J. Magn. Reson.* **88**, 177–185
- 32 Cordier, F. and Grzesiek, S. (1999) Direct observation of hydrogen bonds in proteins by interresidue $^3HJNC'$ scalar couplings. *J. Am. Chem. Soc.* **121**, 1601–1602
- 33 Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302
- 34 Reference deleted
- 35 Brunger, 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. et al. (1998) Acta Crystallogr. Sect. D. *Biol. Crystallogr.* **54**, 905–921
- 36 Nilges, M., Macias, M. J., O'Donoghue, S. I. and Oschkinat, H. (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β -spectrin. *J. Mol. Biol.* **269**, 408–422
- 37 Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. Circumventing problems associated with folding. *FEBS Lett.* **239**, 129–136
- 38 Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M. and Nilges, M. (2003) Refinement of protein structures in explicit solvent. *Proteins* **50**, 496–506
- 39 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
- 40 Sorimachi, K., Jacks, A. J., Le Gal-Coeffet, M. F., Williamson, G., Archer, D. B. and Williamson, M. P. (1996) Solution structure of the granular starch binding domain of glucoamylase from *Aspergillus niger* by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **259**, 970–987
- 41 Abe, A., Tono, T., Sakano, Y. and Kamitori, S. (2004) Complex structures of *Thermoactinomyces vulgaris* R-47 α -amylase 1 with malto-oligosaccharides demonstrate the role of domain N acting as a starch-binding domain. *J. Mol. Biol.* **335**, 811–822
- 42 Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A. and Law, V. (2006) A structural and functional analysis of α -glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *J. Biol. Chem.* **281**, 587–598
- 43 Mikami, B., Iwamoto, H., Malle, D., Yoon, H. J., Demirkan-Sarikaya, E., Mezaki, Y. and Katsuya, Y. (2006) Crystal structure of pullulanase: evidence for parallel binding of oligosaccharides in the active site. *J. Mol. Biol.* **359**, 690–707
- 44 Grzesiek, S., Bax, A., Clore, G. M., Gronenborn, A. M., Hu, J. S., Kaufman, J., Palmer, I., Stahl, S. J. and Wingfield, P. T. (1996) The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. *Nat. Struct. Biol.* **3**, 340–345
- 45 Garrett, D. S., Seok, Y. J., Peterkofsky, A., Clore, G. M. and Gronenborn, A. M. (1997) Identification by NMR of the binding surface for the histidine-containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the *Escherichia coli* phosphotransferase system. *Biochemistry* **36**, 4393–4398
- 46 Dehner, A., Furrer, J., Richter, K., Schuster, I., Buchner, J. and Kessler, H. (2003) NMR chemical shift perturbation study of the N-terminal domain of Hsp90 upon binding of ADP, AMP-PNP, geldanamycin, and radicicol. *ChemBioChem* **4**, 870–877
- 47 Lutgen, H., Robelek, R., Muhlberger, R., Diercks, T., Schuster, S. C., Kohler, P., Kessler, H., Bacher, A. and Richter, G. (2002) Transcriptional regulation by antitermination. Interaction of RNA with NusB protein and NusB/NusE protein complex of *Escherichia coli*. *J. Mol. Biol.* **316**, 875–885
- 48 Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662
- 49 Moseley, H. N., Sahota, G. and Montelione, G. T. (2004) Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. *J. Biomol. NMR* **28**, 341–355
- 50 Rocchia, W., Alexov, D. and Honig, B. (2001) Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. *J. Phys. Chem. B* **105**, 6507–6514
- 51 Machovic, M., Svensson, B., MacGregor, E. A. and Janecek, S. (2005) A new clan of CBM families based on bioinformatics of starch-binding domains from families CBM20 and CBM21. *FEBS J.* **272**, 5497–5513
- 52 Seidel, R. D., Amor, J. C., Kahn, R. A. and Prestegard, J. H. (2004) Structural perturbations in human ADP ribosylation factor-1 accompanying the binding of phosphatidylinositides. *Biochemistry* **43**, 15393–15403
- 53 Williamson, M. P., Le Gal-Coeffet, M. F., Sorimachi, K., Furniss, C. S., Archer, D. B. and Williamson, G. (1997) Function of conserved tryptophans in the *Aspergillus niger* glucoamylase 1 starch binding domain. *Biochemistry* **36**, 7535–7539
- 54 Panyi, T., Szabo, L., Nagy, T., Orosz, L., Simpson, P. J., Williamson, M. P. and Gilbert, H. J. (2000) Trp22, Trp24, and Tyr8 play a pivotal role in the binding of the family 10 cellulose-binding module from *Pseudomonas xylanase* A to insoluble ligands. *Biochemistry* **39**, 985–991

- 55 Xie, H., Bolam, D. N., Nagy, T., Szabo, L., Cooper, A., Simpson, P. J., Lakey, J. H., Williamson, M. P. and Gilbert, H. J. (2001) Role of hydrogen bonding in the interaction between a xylan binding module and xylan. *Biochemistry* **40**, 5700–5707
- 56 Guan, L., Hu, Y. and Kaback, H. R. (2003) Aromatic stacking in the sugar binding site of the lactose permease. *Biochemistry* **42**, 1377–1382
- 57 Pell, G., Williamson, M. P., Walters, C., Du, H., Gilbert, H. J. and Bolam, D. N. (2003) Importance of hydrophobic and polar residues in ligand binding in the family 15 carbohydrate-binding module from *Cellvibrio japonicus* Xyn10C. *Biochemistry* **42**, 9316–9323
- 58 Goto, M., Semimaru, T., Furukawa, K. and Hayashida, S. (1994) Analysis of the raw starch-binding domain by mutation of a glucoamylase from *Aspergillus awamori* var. *kawachi* expressed in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **60**, 3926–3930
- 59 Katsuya, Y., Mezaki, Y., Kubota, M. and Matsuura, Y. (1998) Three-dimensional structure of *Pseudomonas isoamylase* at 2.2 Å resolution. *J. Mol. Biol.* **281**, 885–897
- 60 Souchet, M., Legave, M., Jullian, N., Bertrand, H. O., Bril, A. and Berrebi-Bertrand, I. (1999) Structure of the human glycogen-associated protein phosphatase 1 regulatory subunit hGM: homology modeling revealed an $(\alpha/\beta)_8$ -barrel-like fold in the multidomain protein. *Protein Sci.* **8**, 2570–2579
-

Received 30 August 2006/13 November 2006; accepted 22 November 2006

Published as BJ Immediate Publication 22 November 2006, doi:10.1042/BJ20061312