The family 21 carbohydrate-binding module of glucoamylase from *Rhizopus oryzae* consists of two sites playing distinct roles in ligand binding

Wei-I CHOU*, Tun-Wen PAI†, Shi-Hwei LIU*, Bor-Kai HSIUNG* and Margaret D.-T. CHANG*¹

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

†Department of Computer Science, National Taiwan Ocean University, No. 2, Pei Ning Rd, Keelung, Taiwan 20224, Republic of China

The starch-hydrolysing enzyme GA (glucoamylase) from *Rhizopus oryzae* is a commonly used glycoside hydrolase in industry. It consists of a C-terminal catalytic domain and an N-terminal starch-binding domain, which belong to the CBM21 (carbohydrate-binding module, family 21). In the present study, a molecular model of CBM21 from *R. oryzae* GA (RoGACBM21) was constructed according to PSSC (progressive secondary structure correlation), modified structure-based sequence alignment, and site-directed mutagenesis was used to identify and characterize potential ligand-binding sites. Our model suggests that RoGACBM21 contains two ligand-binding sites, with Tyr³² and Tyr⁶⁷ grouped into site I, and Trp⁴⁷, Tyr⁸³ and Tyr⁹³ grouped into site II. The involvement of these aromatic residues has been validated using chemical modification, UV difference spectroscopy studies, and both qualitative and quantitative binding assays on a series of RoGACBM21 mutants. Our results further reveal that

binding sites I and II play distinct roles in ligand binding, the former not only is involved in binding insoluble starch, but also facilitates the binding of RoGACBM21 to long-chain soluble polysaccharides, whereas the latter serves as the major binding site mediating the binding of both soluble polysaccharide and insoluble ligands. In the present study we have for the first time demonstrated that the key ligand-binding residues of RoGACBM21 can be identified and characterized by a combination of novel bioinformatics methodologies in the absence of resolved three-dimensional structural information.

Key words: binding affinity, carbohydrate-binding module (CBM), glucoamylase (GA), homology modelling, progressive secondary structure correlation (PSSC), starch-binding domain (SBD).

INTRODUCTION

GA (glucoamylase) (1,4- α -D-glucan glucohydrolase, EC 3.2.1.3) is an exo-acting GH that catalyses the release of β -D-glucose from the non-reducing ends of starch and related substrates by hydrolysing α -1,4- and α -1,6-glycosidic linkages using a single active site [1,2]. In industrial utilization, the conversion of starch into glucose syrups by GA from *Rhizopus* or *Aspergillus* is well established [1,3].

GA is a modular protein consisting of a catalytic domain, classified as a member of the GH15 (glycoside hydrolase family 15) [4], connected to an SBD (starch-binding domain) by an O-glycosylated linker. Most of the known SBDs found in GA are located at the C-terminus, but the GAs from *Rhizopus oryzae* [5,6], *Arxula adenivorans* [7] and *Mucor circinelloides* [8] contain N-terminal SBDs. These C-terminal and N-terminal SBDs are categorized into CBM20 (carbohydrate-binding module family 20) and CBM21 respectively.

CBMs are currently classified into 43 families on the basis of amino acid sequence similarity, and exhibit discrepancy in ligand specificity (for detailed information, see <http://www.cazy.org/CAZY/> [9,10]). CBMs mediate the interaction of GHs with their substrates by carrying catalytic domains on the surface of insoluble polysaccharides [11]. Both proteolytic excision and truncation of CBMs from GHs cause a significant decrease in the enzyme's activity towards insoluble but not soluble polysaccharide substrates [10,12,13].

Starch-binding CBMs are grouped into six different families (CBM20, CBM21, CBM25, CBM26, CBM34 and CBM41) [14–

19], among which the structural information for CBM20, CBM25, CBM26 and CBM34 [14, 20–26] is found in the RCSB Protein Data Bank. These structures all contain the characteristic feature termed the β -sandwich fold (comprising two β -sheets) [27]. The studies on the structure and function of CBM20 are the most extensive and detailed [17]; however, only one example of starch-binding function has been demonstrated in CBM21 [28]. Some CBM21 members bind to insoluble storage polysaccharides such as starch and glycogen [29], whereas most of the others are hypothetical proteins derived from genomic annotation [30].

The most recent evolutionary study on CBM20 and CBM21 proposes to group these two SBD families into a new CBM clan [31]. It also indicates that the putative fold of CBM21, identified by the three dimensional PSSC (progressive secondary structure correlation) fold recognition server [32], is similar to that of CBM20. CBM21 and CBM20 are predicted to have similar secondary and tertiary structures and the same role in enhancing the activity of hydrolysing insoluble starch, even though their sequence identity is fairly low. It is postulated that the CBM21 from *R. oryzae* GA (RoGACBM21), as well as CBM20 from *Aspergillus niger* GA (AnGACBM20), should belong with the members of Type B glycan chain-binding CBMs [27] that have similar structural topology. The location and orientation of the aromatic residues in the binding sites of Type B CBMs are known to be the primary factors in ligand specificity and affinity [33,34]. Another interaction, direct hydrogen bonding, is also an important factor in CBM-ligand recognition since the hydroxy groups of carbohydrates are capable of forming hydrogen bonds with polar

Abbreviations used: Aa, *Arxula adenivorans*; An, *Aspergillus niger*; BCA, bichinchonic acid; CBM21, carbohydrate-binding module, where 21 is family 21 etc; GA, glucoamylase; GH15, glycoside hydrolase, where 15 is family 15 etc; HsPP, *Homo sapiens* protein phosphatase; LkA, *Lipomyces kononenkoae* α -amylase; Mc, *Mucor circinelloides*; NBS, N-bromosuccinimide; PSSC, progressive secondary structure correlation; Ro, *Rhizopus oryzae*; SBD, starch-binding domain; TNM, tetranitromethane.

¹ To whom correspondence should be addressed (email dtchang@life.nthu.edu.tw).

residues (mainly Glx and Asx) in the binding sites of CBMs [35,36].

The structural information for CBM20 provided by NMR spectroscopic and X-ray crystallographic studies has been utilized to study the biological functions of the members of this family [25,26]; however, no three-dimensional structure of any CBM21 protein is currently available in any database. Furthermore, molecular modelling of intact *RoGACBM21* using *AnGACBM20* as the template cannot be directly processed in the SWISS MODEL protein-modelling server because of the low sequence identity (13.5%) between *RoGACBM21* and *AnGACBM20*. To solve this problem, the PSSC algorithm and structure-based sequence alignment were used in this study to compare the representative CBM21 members and *AnGACBM20* that share similar topologies. The result of the modified structure-based sequence alignment indicates that the positions of aromatic residues and the lengths of β -strands are aligned much better than in conventional alignment. The molecular model of *RoGACBM21* was constructed according to the result of structure-based sequence alignment and was validated experimentally by site-directed mutagenesis, chemical modification, UV difference spectroscopy, and both qualitative and quantitative binding assays. In addition, the roles of the aromatic residues were also characterized. The results of the present study show that *RoGACBM21* has an all- β -strand structure containing two ligand-binding sites, similar to that found in *AnGACBM20*. Interestingly, unlike *AnGACBM20*, *RoGACBM21* has distinct ligand specificity for each binding site.

EXPERIMENTAL

Structure-based multiple sequence alignment

Secondary-structure prediction was performed at the Jpred server (<http://www.compbio.dundee.ac.uk/~www-jpred/> [37]) and the NPSA (Network Protein Sequence Analysis) server (<http://npsa-pbil.ibcp.fr> [38]). The Jpred and NPSA servers provide the consensus results of six (NNSP, DSC, PREDATOR, MULPRED, ZPRED and PHD) and twelve (SOPM, SOPMA, HNN, MLRC, DPM, DSC, GORI, GORII, GORIV, PHD, PREDATOR and SIMPA96) different methods respectively. The results of secondary structure prediction were used for multiple sequence alignment of the amylolytic CBM21 SBDs from *RoGACBM21*, *A. adenivorans* GA (*AaGACBM21*), *M. circinelloides* GA (*McGACBM21*) and *Lipomyces kononenkoae* α -amylase (*Lk α ACBM21*), the non-amylolytic CBM21 from *Homo sapiens* protein phosphatase (*HsPPCBM21*) and *AnGACBM20* using the ClustalX program [39] with the option of the secondary structure profile alignment.

PSSC algorithm

In the present study, we developed a PSSC algorithm, which was based on combinatorial informatics for predicting secondary structures and functionally important residues from related sequences of protein families. PSSC comprises three major steps: (i) obtaining multiple secondary structures using existing prediction systems to assign positions and boundaries of α -helices and β -strands, (ii) transforming each of the predicted secondary structures into a new sequence representation by symbolizing the α -helices and β -strands, as well as labelling selected aromatic residues, tyrosine, tryptophan and phenylalanine, pre-defined as constrained residues, and (iii) performing progressive correlation operations on one protein sequence with a resolved three-dimensional structure as the template. Among all predicted secondary structures, the one with the highest score represents the best candidate in which to identify the key residues corresponding

to those in the template structure. Each annotated segment of the α -helix or β -strand was simplified by a symbol ' α ' and ' β ' respectively, and the loop structures were replaced with specified constrained residues or a symbol 'N' when no aromatic residue could be allocated. In the last step, the correlation operations were performed to provide the best alignment with respect to the location of secondary structures and constrained aromatic residues. Different weightings for matching scores were assigned to enhance the secondary structure relationship. A score of 2 was assigned to any matched symbol ' α ' and ' β ', a score of 1 was given to a matched 'Y', 'W', 'F' and 'N', and a score of -1 was assigned as a mismatch penalty. The predicted secondary structures with the highest scores were selected for identification of key residues through further correlation alignment.

Model construction

A molecular model of *RoGACBM21* was generated by the protein-modelling server SWISS MODEL [40–42] using the NMR structure of *AnGACBM20* (PDB code 1AC0 [25]) as the template. Based on the modified sequence alignment, residues 1–13 were deleted from the *RoGACBM21* model structure, as there were no corresponding residues in *AnGACBM20*. Energy minimization of the model was performed with the GROMOS96 implementation of the DeepView program [41]. Molecular graphics were performed using WebLab ViewerLite.

Protein expression and purification

The DNA fragment encoding the SBD of *R. oryzae* GA from Ala¹ to Thr¹⁰⁶ was amplified by PCR using the forward primer 5'-CATATGGCAAGTATTCCTAGCAGT-3' and the reverse primer 5'-CTCGAGTGTAGATACTTGGT-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-*RoGACBM21*. *Escherichia coli* BL21-Gold (DE3) cells (Novagen), transformed with pET-*RoGACBM21*, were grown in Luria–Bertani medium containing 100 $\mu\text{g} \cdot \text{ml}^{-1}$ ampicillin, and incubated at 37°C with shaking at 250 g until the attenuation (D_{600}) reached 0.6. The synthesis of proteins was induced by the addition of IPTG (isopropyl β -D-thiogalactoside) to a final concentration of 400 μM , the incubation was continued at 20°C for 16 h. The cells were harvested by centrifugation at 3700 g for 15 min at 4°C, and the resultant pellet that was resuspended in 20 ml of Tris/HCl buffer (50 mM, pH 7.9) was then homogenized (EmulsiFlex-C5 homogenizer). The cell debris was removed by centrifugation at 16000 g for 15 min at 4°C, and the supernatant was subjected to purification by His-Bind[®] affinity column chromatography (Novagen). The purified *RoGACBM21* was dialysed against sodium acetate buffer (50 mM, pH 5.5) using an Amicon stirred-cell concentrator (Millipore) equipped with a PM-10 membrane (10 kDa cutoff). Protein concentrations were assayed by a BCA (bicinchoninic acid) reagent kit (Pierce).

Site-directed mutagenesis

All *RoGACBM21* mutants were generated using the PCR-based QuikChange site-directed mutagenesis method (Stratagene) with pET-*RoGACBM21* as the template, two complementary primers containing the desired mutation, and Pfu Turbo DNA polymerase (Stratagene). The sequence of each mutant plasmid was verified to confirm that additional PCR-induced mutations had not been introduced. All constructs were transformed into competent *E. coli* BL21-Gold (DE3) for protein expression.

CD

CD spectra were recorded on an Aviv CD spectrometer (model 202) equipped with a 450-W xenon arc lamp. Far-UV spectral analysis at 200–260 nm was performed in a rectangular quartz cuvette with a 0.1 cm path length at 25 and 96 °C at a scan speed of 4 nm · s⁻¹ and a bandwidth of 0.5 nm. Each spectrum obtained was an average from three consecutive scans and was corrected by subtracting the buffer spectrum.

UV difference spectra of wild-type and mutant RoGACBM21

All spectra were obtained using a UV/visible spectrophotometer (Hitachi U-3310) at 270–300 nm with a bandwidth of 0.5 nm. Spectra of 30 µM wild-type or mutant RoGACBM21 were recorded in 50 mM sodium acetate, pH 5.5, at 25 °C in the absence or presence of 500 µM β-cyclodextrin (Sigma). Each difference spectrum, calculated by subtracting the protein-only spectrum from protein–ligand spectrum, was used for comparison with the difference spectra of model compounds *N*-acetyltryptophan (50 µM) (Sigma) and *N*-acetyltyrosine (100 µM) (Sigma), as perturbed in 20 % DMSO.

Chemical modification of tryptophan residues with NBS (*N*-bromosuccinimide)

The oxidation of tryptophan residues was performed by titrating native wild-type and W47A RoGACBM21 protein (30 µM, in 50 mM sodium acetate, pH 5.5), native wild-type protein in the presence of 500 µM β-cyclodextrin or of denaturant (8 M urea) with NBS (1 mM, freshly prepared) (Sigma) in a rectangular quartz cuvette with a 1 cm path length at 25 °C. At each titration point, the mixture was incubated for 3 min, and the absorbance at 280 nm was recorded using a UV/visible spectrophotometer. The titration process was continued until the decreasing trend of absorbance at 280 nm stagnated or started to rise. The number of tryptophan residues oxidized by NBS was calculated using the absorbance method [43].

Chemical modification of tyrosine residues with TNM (tetranitromethane)

The nitration of tyrosine residues was performed using TNM, as described by Sokolovsky et al. [44]. RoGACBM21 (10 µM, in 100 mM Tris/HCl, pH 8.0) was added into TNM (1 mM, in 95 % ethanol) at 25 °C for 180 min, and the absorbance at 428 nm was monitored using a UV/visible spectrophotometer. The number of tyrosine residues oxidized by TNM was calculated using the molar absorption coefficient value of 4100 M⁻¹ · cm⁻¹ for 3-nitrotyrosine [44].

Qualitative measurement of binding to insoluble starch

Wild-type or mutant RoGACBM21 (16 µM, 100 µl) in 50 mM sodium acetate, pH 5.5, was each mixed with 0.1 mg of insoluble starch (Sigma), and the mixture was incubated at 25 °C with gentle stirring for 5 h, after which the insoluble starch was pelleted by centrifugation at 13 000 *g* for 2 min at 4 °C. The pellet (the bound fraction) and supernatant (the unbound fraction) were analysed by SDS/PAGE using a 15 % gel. To ensure that no precipitation occurred during the assay period, protein with no insoluble starch was included as a control.

Quantitative measurement of binding to insoluble starch

The starch-binding isotherm was analysed as a saturation binding assay. Wild-type and mutant RoGACBM21 (100 µl, in 50 mM

sodium acetate, pH 5.5) were each mixed with 0.1 mg of pre-washed insoluble starch and incubated at 25 °C with gentle stirring for 16 h. After centrifugation at 16 000 *g* for 10 min at 4 °C, the protein concentration of the supernatant was determined using the BCA assay, and the amount of bound protein was calculated from the difference between the initial and unbound protein concentrations. The dissociation constant (K_d) and the maximal amount of bound protein (B_{max}) were determined by fitting to the non-linear regression of the binding isotherms using a standard single-site binding model.

Quantitative measurement of binding to soluble polysaccharides

Fluorescence spectrophotometry of the binding of wild-type or mutant RoGACBM21 to soluble polysaccharides (Sigma) was recorded by measuring changes in the intrinsic protein fluorescence intensity. Experiments were performed in 50 mM sodium acetate, pH 5.5, at 25 °C using a PerkinElmer LS-55 spectrophotometer. Circular and linear carbohydrates (2–20 mM) were titrated into RoGACBM21 (10 µM, 2 ml), and the fluorescence-emission spectrum was monitored at 350 nm with a fixed excitation at 280 nm. The relative changes in fluorescence intensity were plotted against ligand concentration, and the data were fitted to a simulated curve using the appropriate equation for a single binding site.

Data analysis

Data analyses of kinetic and affinity parameters were performed using GraphPad Prism (GraphPad Software).

RESULTS AND DISCUSSION

Construction of the RoGACBM21 model

Six CBM families, CBM20, CBM21, CBM25, CBM26, CBM34, and CBM41, contain SBDs, but currently only the structures of CBM20, CBM25, CBM26 and CBM34 have been elucidated [14,20–26]. All of the resolved three-dimensional structures share an immunoglobulin-like topology (CATH code 2.60.40) [45], suggesting that all SBDs of GA possess similar three-dimensional structures, even though they do not have high sequence identity [31].

Because the CBM21 sequence does not have a high level of identity with any previously solved CBM20 structure, a novel structure-based multiple sequence alignment strategy was used to compare proteins that share similar topology but are dissimilar in sequence. The secondary structures of the SBDs used in this study were initially predicted using the consensus secondary-structure prediction servers, Jpred and NPSA to construct a consensus result using a combination of a number of secondary-structure prediction methods.

Figure 1(A) presents the results of the structure-based multiple sequence alignment for CBM20 and CBM21. The consensus secondary structures for all CBM21 SBDs except AaGACBM21 revealed eight β-strands, consistent with the structural characteristics of CBM20, whereas the seventh β-strand in AaGACBM21 was replaced by an α-helix. This phenomenon has also been mentioned by Machovic et al. [31]. Most of the secondary structural elements of both CBM families appear to be located in corresponding positions without considering the length of each β-strand; however, no consensus aromatic residues between CBM20 and CBM21 were revealed. In an attempt to improve the present alignment for identifying potential ligand-binding residues in RoGACBM21, the PSSC algorithm was used with structure-based multiple sequence alignments. We hypothesized that strongly correlated results could be obtained not only from

Figure 1 Comparison of the structure-based multiple sequence alignment of CBM20 and CBM21

(A) Structure-based multiple sequence alignment with the amylolytic CBM21 SBDs from *RoGACBM21*, *AaGACBM21*, *McGACBM21* and *LkαACBM21*, the non-amylolytic *HsPPCBM21*, and *AnGACBM20*. The white, black and grey segments represent the predicted secondary-structural element loop, β -strand and α -helix respectively. Y, W and F in black indicate tyrosine, tryptophan and phenylalanine in the loop region respectively; Y, W and F in white represent the same aromatic residues in β -strands. (B) Structure-based sequence alignment using the PSSC algorithm with the schematized sequences of *RoGACBM21* and *AnGACBM20* marked above and below the Figure respectively. β , β -Strand; N, no aromatic residue in the loop region. The key ligand-binding residues in *AnGACBM20* are denoted by a double underline.

Figure 2 Molecular modelling of *RoGACBM21* using structure-based sequence alignment

(A) The locations of the β -strands in *RoGACBM21* and *AnGACBM20*, represented as colour-coded arrows, are indicated above and below the sequences respectively. The ligand-binding sites in *AnGACBM20* are underlined (binding site I, single underline; binding site II, double underline). The aromatic residues predicted to be involved in carbohydrate binding in *RoGACBM21* are labelled in red. (B) The modelling structure of *RoGACBM21* (left panel) and the template structure of *AnGACBM20* (right panel, PDB code 1AC0) are shown as ribbon diagrams. The polypeptide sequence from Ala¹ to Ser¹³ forming the first β -strand of *RoGACBM21* is not shown in the model. The ligands (β -cyclodextrin) and the side-chains of the characterized ligand-binding sites in *AnGACBM20*, as well as the potential binding sites of *RoGACBM21* are also depicted. (C) The topology of secondary-structural directionality, and the organization of *RoGACBM21* and *AnGACBM20* are shown with colour-coded arrows corresponding to those in (A) and (B).

relative α -helical and β -stranded structures, but also from key residues on loop regions that participate in ligand binding to maintain functionality, similar to the template. As shown in Figure 1(B), the secondary-structural elements were well-aligned in terms of relative position and length, and eight aromatic residues in RoGACBM21 were found to be conserved with the corresponding residues in AnGACBM20. Compared with the recently published alignment data of all CBM21 and CBM20 members [31], PSSC shows that the second to the eighth β -strands of RoGACBM21 are sequentially aligned with the first to the seventh β -strands of AnGACBM20, instead of inserting a large gap in the region between the second and third β -strands of RoGACBM21 during multiple sequence alignment. Our results revealed that Tyr³², Trp⁷⁰ and Tyr⁹³ (RoGACBM21 numbering) are absolutely conserved in all amylolytic CBM21 SBDs, including two α -amylases and three GAs. Interestingly, these three aromatic residues in the equivalent positions in the non-amylolytic CBM21 are all replaced with phenylalanine. In addition, it was found that Trp⁴⁷, Tyr⁵⁸ and Tyr⁹⁴ were better conserved in CBM21 from GAs and non-amylolytic enzymes than in CBM21 SBDs from α -amylases.

As shown in Figure 2(A), several potential ligand-binding residues (Tyr³², Trp⁴⁷, Tyr⁵⁸, Tyr⁶⁷, Tyr⁸³, Tyr⁹³ and Tyr⁹⁴) in RoGACBM21, that are located in equivalent or vicinal positions in AnGACBM20 are identified and labelled in red. The molecular model of RoGACBM21 (Figure 2B, left panel) was generated from the residues spanning positions 14–106, as aligned with AnGACBM20. The root mean square deviation of the superimposed backbone atoms (356 atoms) was 1.38 Å (1 Å = 0.1 nm), indicating that the model fitted the template structure very well. Judging from the AnGACBM20 structure with two ligand-binding sites (Figure 2B, right panel), we hypothesized that the ligand-binding residues in RoGACBM21 could also be grouped into two sites, namely site I (Tyr³², Tyr⁵⁸ and Tyr⁶⁷) and site II (Trp⁴⁷, Tyr⁸³, Tyr⁹³ and Tyr⁹⁴), with Tyr⁵⁸ and Tyr⁹⁴ located on the side chains oriented towards the inside of the RoGACBM21 structure, and not involved in ligand binding.

The RoGACBM21 model includes a three-stranded β -sheet and a five-stranded β -sheet (Figure 2C, left panel). These two β -sheets consist of anti-parallel β -strands with the topological organizations (1 $\downarrow$ 2 $\uparrow$ 5 $\downarrow$) and (4 $\uparrow$ 3 $\downarrow$ 6 $\uparrow$ 7 $\downarrow$ 8 $\downarrow$). β -Strands 2–8 of RoGACBM21 could be superimposed on to β -strands 1–7 of AnGACBM20 (Figure 2C, right panel), and the first strand of RoGACBM21 could possibly be superimposed on to the eighth β -strand of AnGACBM20, although in the reverse direction.

Validation of the RoGACBM21 Model

All- β -strand secondary structure

The CD spectrum of the native state RoGACBM21 (Figure 3) had a trough at 215 nm, typical for a protein containing only β -strands, suggesting that RoGACBM21 does not contain a helical structure, and that both β -strand and random coil structures significantly contribute to the overall conformation. This observation is in agreement with the results of the secondary structure prediction and the shared all- β -strand structure of the starch-binding CBMs. The spectrum also contained a peak near 230 nm, an unusual feature that has been observed in some CBDs in which disulphide bonds and aromatic side chains of phenylalanine, tyrosine and tryptophan may contribute to spectral peaks with positive ellipticity [46,47]. RoGACBM21 contains 18 aromatic residues and no cysteine residue (Figure 2A), suggesting that the peak of positive ellipticity near 230 nm may be a consensus sequence of aromatic residues.

Figure 3 Determination of the secondary structure of RoGACBM21 by far-UV CD

The CD spectrum of the native RoGACBM21 in 50 mM sodium acetate (pH 5.5) at 25 °C was determined. Abbreviation: m.r.w., mean residue weight.

Distinct roles of the two tryptophan residues in RoGACBM21

NBS was employed to further evaluate the number of tryptophan residues exposed on the surface of RoGACBM21. NBS modification of the tryptophan residues in RoGACBM21 was monitored by recording the decrease in absorbance at 280 nm (Figure 4A). The decrease in A_{280} of native RoGACBM21 began to rise at an NBS/protein ratio of 2.5:1. This NBS-induced rise in absorbance at 280 nm is a phenomenon typically observed following the oxidation of tyrosine residues [48]. However, no increase in the A_{280} of denatured RoGACBM21 was observed since none of the tyrosine residues in the urea solution could be modified by NBS [44]. As shown in Figure 4(A), the number of modified tryptophan residues in native and denatured RoGACBM21 was calculated to be 1.6 and 2.0 respectively. The only two tryptophan residues residing in RoGACBM21 could be modified under native and denaturing conditions, indicating that both Trp⁴⁷ and Trp⁷⁰ may be exposed on the surface of the protein. To determine whether two tryptophan residues are located within the ligand-binding site(s), NBS oxidations were carried out in the presence of the ligand, β -cyclodextrin. Pre-incubation of native RoGACBM21 with β -cyclodextrin resulted in the apparent protection of one tryptophan residue from oxidation (Figure 4A), suggesting that only one tryptophan residue is involved in β -cyclodextrin binding.

To further probe which tryptophan residue is located in the ligand-binding site, both Trp⁴⁷ and Trp⁷⁰ were individually mutated to alanine. The W47A mutant could be expressed in and purified from *E. coli*; however, the solubility of the W70A mutant was very poor under all expression conditions tested, suggesting that Trp⁷⁰ is quite important for the correct folding of RoGACBM21. Our result is in good agreement with the observation that the highly conserved Trp⁵⁶³ (AnGACBM20 numbering) in both amylolytic CBM20 and CBM21 SBDs may play a structural role [49].

The native W47A mutant was subjected to treatment with NBS, and the result was consistent with that of the wild-type RoGACBM21 tested in the presence of β -cyclodextrin (Figure 4A), strongly suggesting that Trp⁴⁷ is the key ligand-binding residue. This result was also confirmed by UV difference spectra for wild-type and mutant RoGACBM21 (Figure 4B),

Figure 4 Chemical modifications and UV difference spectra of *RoGACBM21* derivatives

(A) Native wild-type *RoGACBM21* in the absence (■) or presence (□) of 500 μ M β -cyclodextrin; denatured wild-type *RoGACBM21* (▲) and native W47A in the absence (●) or presence (○) of 500 μ M β -cyclodextrin were titrated with NBS. The data were normalized to adjust the original absorbance of each reaction to 1.0. (B) The difference spectra of *N*-acetyltryptophan (50 μ M) (a) and *N*-acetyltyrosine (100 μ M) (b) perturbed with 20% DMSO were used for comparison. Difference spectra of wild-type recombinant *RoGACBM21* (c), Y32A (d), W47A (e), Y16A (f) and Y86A (g) perturbed with β -cyclodextrin were normalized to an equivalent protein concentration. (C) Wild-type *RoGACBM21* (10 μ M, in 100 mM Tris/HCl, pH 8.0) was titrated with 100-fold excess TNM at 25 °C. The data reported are the means from three independent titrations.

which were perturbed with 500 μ M β -cyclodextrin and then compared with those of the model compounds *N*-acetyltryptophan and *N*-acetyltyrosine perturbed with 20% DMSO. The spectrum of wild-type *RoGACBM21* (Figure 4B, line c), with three peaks near 277, 285 and 293 nm resembled the tryptophan and tyrosine perturbation spectra (Figure 4B, lines a and b), suggesting that the perturbed tryptophan and tyrosine residue(s) in *RoGACBM21* are present in the ligand-binding site(s). The difference spectrum of W47A (Figure 4B, line e) was diminished and virtually identical with the baseline scan, revealing that Trp⁴⁷ is involved in soluble ligand binding. These data demonstrate that Trp⁴⁷ plays a critical role in binding to soluble ligands and that Trp⁷⁰ is essential for maintaining the structure.

Identification of key tyrosine residues for ligand binding

The tyrosine residues in *RoGACBM21* were nitrated with TNM and the numbers of tyrosine residues exposed on the surface were calculated using the molar absorption coefficient of the nitrated tyrosine (3-nitrotyrosine) (Figure 4C). More than six of the 12 tyrosine residues were nitrated by TNM, suggesting that at least 50% of the tyrosines were exposed on the surface and thus may contribute to a potential ligand-binding site(s). To further investigate the crucial tyrosine residues located in *RoGACBM21*, each were individually replaced with alanine. Qualitative binding assays were carried out to probe the number and location of potential ligand-binding residues in wild-type *RoGACBM21*, the tryptophan mutant (W47A) and tyrosine mutants (Y32A, Y58A, Y67A, Y83A, Y93A and Y94A). As shown in Figure 5, substitutions of alanine for Tyr³², Trp⁴⁷, Tyr⁶⁷, Tyr⁸³ and Tyr⁹³ significantly decreased the binding to insoluble starch; as expected no change was observed in mutants Y58A and Y94A. These results validated our assumptions based simply on the three-dimensional model of *RoGACBM21*. Nevertheless, the solubility of Y67A, Y83A and Y93A was much lower than that of the other mutants, indicating that Tyr⁶⁷, Tyr⁸³ and Tyr⁹³

Figure 5 Qualitative binding of *RoGACBM21* derivatives to insoluble starch, as analysed by SDS/PAGE

Purified wild-type (WT) and Y16A, Y32A, W47A, Y58A, Y67A, Y83A, Y86A, Y93A and Y94A mutants of *RoGACBM21* were incubated with insoluble starch, and then pelleted by centrifugation. S, protein not bound to insoluble polysaccharide; P, protein bound to insoluble polysaccharide.

are also important for stabilizing the *RoGACBM21* structure. Accordingly, no soluble double-mutant protein involving these residues was obtained. Therefore subsequent studies focused on Tyr³² and Trp⁴⁷.

The difference UV spectra of mutant Y32A (Figure 4B, line d) and wild-type *RoGACBM21* did not differ significantly, indicating that binding site I is not involved in β -cyclodextrin binding. It should be noted that putative binding residues Tyr¹⁶ and Tyr⁸⁶ that were inferred from direct sequence alignment are not involved in the binding to either soluble (Figure 4B, lines f and g) or insoluble ligand (Figure 5).

Two binding sites with distinct ligand specificities

Equilibrium starch-binding isotherm data for purified *RoGACBM21* were fitted to non-linear regression curves (Figure 6). The binding parameters, K_d and B_{max} , for *RoGACBM21*, as determined by a two-parameter model for single-site saturation binding were $1.43 \pm 0.14 \mu$ M and $41.14 \pm 1.05 \mu$ mol $\cdot$ g⁻¹ respectively (Table 1). However, the K_d for the wild-type

Figure 6 Binding of RoGACBM21 derivatives to insoluble starch, as analysed by depletion isotherms

Starch-binding assays of RoGACBM21 derivatives were performed with insoluble corn starch at different protein concentrations. Binding isotherm (left panel) and Scatchard analysis (right panel) of wild-type RoGACBM21 (■), Y32A (▲), W47A (●) and Y32A/W47A (□). The data reported are the means for three independent experiments.

Table 1 Affinity of RoGACBM21 derivatives for insoluble starch determined by depletion isotherms

K_d and B_{max} values were derived from saturation binding experiments as described in the Experimental section. Values are the means $\pm$ S.D. for three independent experiments.

Protein	K_d (μ M)	B_{max} (μ mol $\cdot$ g $^{-1}$)
Wild-type	1.4 ± 0.1	41.1 ± 1.1
Y32A	0.8 ± 0.1	23.4 ± 0.7
W47A	4.4 ± 0.6	23.0 ± 1.3
Y32A/W47A	25.0 ± 2.9	5.1 ± 0.3

RoGACBM21 is an apparent constant representing a weighted average over two binding sites. Compared with AnGACBM20, the dissociation constants for the engineered AnGACBM20 ($3.2 \pm 0.9 \mu$ M) and proteolytically produced AnGACBM20 ($12.7 \pm 0.5 \mu$ M) [50,51] were approx. 2- and 10-fold higher than that of RoGACBM21 respectively. Furthermore, the B_{max} values for the engineered AnGACBM20 (0.56μ mol $\cdot$ g $^{-1}$) and proteolytically produced AnGACBM20 ($1.08 \pm 0.02 \mu$ mol $\cdot$ g $^{-1}$) were respectively 70- and 40-fold lower [50,51] than that of RoGACBM21. It should be noted that the ligand-binding capacity may be affected by various ratios of amylopectin to amylose, the size and shape of starch granules, and the swelling effect arising from the time that the starch sits in the buffer. The quantitative binding isotherm and Scatchard analysis (Figure 6 and Table 1) revealed that the binding capacity of insoluble starch for Y32A and W47A mutants was decreased to approx. 50% relative to that of the wild-type. The binding affinity of the Y32A and W47A mutants was similar to that of the wild-type protein, and the binding of the Y32A/W47A double mutant to insoluble starch was almost completely abolished.

The data for quantitative binding to soluble ligands is presented in Table 2. For cyclic carbohydrates, the affinity of wild-type RoGACBM21 for β -cyclodextrin ($K_d = 5.1 \pm 0.7 \mu$ M) was slightly stronger than for γ CD ($K_d = 8.3 \pm 1.5 \mu$ M), and the binding to α -cyclodextrin ($K_d > 300 \mu$ M) was relatively poor. For linear carbohydrates, the K_d values of wild-type RoGACBM21 decreased with increasing ligand length. These results indicate that the distance and the orientation between the aromatic side chains residing in the binding site are the major determinants for

Table 2 Affinity of RoGACBM21 derivatives for soluble oligosaccharides determined by fluorescence titration spectra

Values are the means $\pm$ S.D. for three independent titrations. No binding was detected in the case of protein W47A.

Ligands	Protein . . .	K_d (μ M)	
		Wild-type	Y32A
γ CD		8.3 ± 1.5	6.6 ± 0.8
β CD		5.1 ± 0.7	6.0 ± 0.7
α CD		333.0 ± 13.1	286.3 ± 30.5
Maltoheptaose		5.5 ± 1.0	44.7 ± 5.3
Maltohexaose		24.4 ± 1.7	46.4 ± 5.8
Maltopentaose		27.5 ± 1.4	50.2 ± 6.2
Maltotetraose		54.4 ± 4.8	53.3 ± 5.5
Maltotriose		61.9 ± 6.0	58.2 ± 8.1
Isomaltotriose		3.8 ± 0.7	3.9 ± 0.8
NB, no binding detected.			

NB, no binding detected.

ligand binding to RoGACBM21. Therefore the binding to cyclic carbohydrates was dependent upon the angle of the ligand surface, whereas binding to linear carbohydrates was governed by ligand length. For the ligand with an α -1,6-glycosidic linkage, the affinity of RoGACBM21 for isomaltotriose ($K_d = 3.8 \pm 0.7 \mu$ M) was stronger than those of the soluble ligands with α -1,4-glycosidic linkage, suggesting that RoGACBM21 may prefer to bind to the side branch of the starch molecule.

The affinity of the W47A mutant for the soluble carbohydrates used in this study either was too low to determine by fluorescence titration spectra or was substantially affected by the alanine substitution. These findings, together with the results of chemical modification, and qualitative and quantitative binding assays for insoluble starch binding, strongly support the hypothesis that Trp⁴⁷ is indeed the major residue in binding site II that is involved in binding to both soluble and insoluble ligands. The Y32A mutant and wild-type RoGACBM21 show similar affinity for various cyclodextrins, confirming that Tyr³² is not involved in binding to such ligands. In addition, the effect of the length of linear oligosaccharides on the binding to Y32A was relatively insignificant as compared with that of wild-type RoGACBM21, suggesting that binding site I may facilitate the binding of RoGACBM21 to

long-chain linear oligosaccharides. The starch-binding CBM20s have been demonstrated to possess two binding sites with similar binding affinities [49,52]; however, *RoGACBM21* in this case contains two binding sites with distinct binding specificities, indicating that each site plays a different but co-operative role in carbohydrate–protein recognition. Binding site II serves as the major binding site for both soluble and insoluble ligands, and binding site I is involved in assisting *RoGACBM21* in association with long-chain carbohydrates.

Conclusion

In conclusion, our data provide the first functional characterization and binding site identification of the SBD of CBM21. Our molecular model of *RoGACBM21*, based on structure-based sequence alignment, is useful for the design of site-directed mutants and for functional analyses. The model also presents the structure–function rationale for qualitative and quantitative binding data. Our results indicate that the Tyr³² and Trp⁴⁷ are located on the surface of *RoGACBM21*, and that each residue plays a distinct role in ligand binding. Of the two insoluble starch-binding sites in *RoGACBM21*, site I is involved in binding only to long-chain soluble polysaccharides. These properties may be important for delivering the ligand from the SBD to the catalytic domain. The structural resemblance between *RoGACBM21* and *AnGACBM20* suggests that it is feasible to search for the key ligand-binding residues in CBM family members even though these proteins share low sequence identity.

This work was supported by Simpson Biotech Co. and the NSC (National Science Council), Taiwan, R.O.C. (NSC grant numbers 94-2627-B-007-003, 94-2752-B-007-003-PAE and 94-3112-B-007-004-Y). We thank Dr P.-C. Huang, Dr N.E. Marsh, Dr R.-L. Pan, Dr P.-C. Lyu, Dr W.-G. Chou, Dr Y.-K. Lai, Dr H.-Y. Chang, Dr M.-F. Tam and Dr Y.C. Sun for their critical comments, as well as Mr C.-C. Sheu and Mr T.-H. Hsu for their discussions regarding the manuscript.

REFERENCES

- Coutinho, P. M. and Reilly, P. J. (1997) Glucoamylase structural, functional, and evolutionary relationships. *Proteins* **29**, 334–347.
- Sauer, J., Sigurskjold, B. W., Christensen, U., Frandsen, T. P., Mirgorodskaya, E., Harrison, M., Roepstorff, P. and Svensson, B. (2000) Glucoamylase: structure/function relationships, and protein engineering. *Biochim. Biophys. Acta* **1543**, 275–293.
- Chiba, S. (1997) Molecular mechanism in α -glucosidase and glucoamylase. *Biosci. Biotechnol. Biochem.* **61**, 1233–1239.
- Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316.
- Ashikari, T., Nakamura, N., Tanaka, Y., Kiuchi, N., Shibano, Y., Tanaka, T., Amachi, T. and Yoshizumi, H. (1986) Rhizopus raw-starch-degrading glucoamylase: its cloning and expression in yeast. *Agric. Biol. Chem.* **50**, 957–964.
- Tanaka, Y., Ashikari, T., Nakamura, N., Kiuchi, N. and Shibano, Y. (1986) Comparison of amino acid sequences of three glucoamylases and their structure–function relationships. *Agric. Biol. Chem.* **50**, 965–969.
- Bui, D. M., Kunze, I., Förster, S., Wartmann, T., Horstmann, C., Manteuffel, R. and Kunze, G. (1996) Cloning and expression of an *Arxula adenivorans* glucoamylase gene in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **44**, 610–619.
- Houghton-Larsen, J. and Pedersen, P. A. (2003) Cloning and characterisation of a glucoamylase gene (GlaM) from the dimorphic zygomycete *Mucor circinelloides*. *Appl. Microbiol. Biotechnol.* **62**, 210–217.
- Coutinho, P. M. and Henrissat, B. (1999) The modular structure of cellulases and other carbohydrate-active enzymes: an integrated database approach. In *Genetics, Biochemistry and Ecology of Cellulose Degradation* (Ohmiya, K., Hayashi, K., Sakka, K., Kobayashi, Y., Karita, S. and Kimura, T., eds), pp. 15–23. Uni Publishers Co., Tokyo.
- Hall, J., Black, G. W., Ferreira, L. M., Millward-Sadler, S. J., Ali, B. R., Hazlewood, G. P. and Gilbert, H. J. (1995) The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluorescens* subsp. *cellulosa* is important for the efficient hydrolysis of Avicel. *Biochem. J.* **309**, 749–756.
- 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.
- Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J. and 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.
- Ali, M. K., Hayashi, H., Karita, S., Goto, M., Kimura, T., Sakka, K. and Ohmiya, K. (2001) Importance of the carbohydrate-binding module of *Clostridium stercorarium* Xyn10B to xylan hydrolysis. *Biosci. Biotechnol. Biochem.* **65**, 41–47.
- Abe, A., Tonoizuka, 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.
- Janecek, S. and Sevcik, J. (1999) The evolution of starch-binding domain. *FEBS Lett.* **456**, 119–125.
- Janecek, S., Svensson, B. and MacGregor, E. A. (2003) Relation between domain evolution, specificity, and taxonomy of the α -amylase family members containing a C-terminal starch-binding domain. *Eur. J. Biochem.* **270**, 635–645.
- Rodriguez-Sanoja, R., Oviedo, N. and Sanchez, S. (2005) Microbial starch-binding domain. *Curr. Opin. Microbiol.* **8**, 260–267.
- Sumitani, J., Tottori, T., Kawaguchi, T. and Arai, M. (2000) New type of starch-binding domain: the direct repeat motif in the C-terminal region of *Bacillus* sp. no. 195 α -amylase contributes to starch binding and raw starch degrading. *Biochem. J.* **350**, 477–484.
- Svensson, B., Jespersen, H., Sierks, M. R. and MacGregor, E. A. (1989) Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. *Biochem. J.* **264**, 309–311.
- 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.
- Kamitori, S., Kondo, S., Okuyama, K., Yokota, T., Shimura, Y., Tonoizuka, T. and Sakano, Y. (1999) Crystal structure of *Thermoactinomyces vulgaris* R-47 α -amylase II (TVAIL) hydrolyzing cyclodextrins and pullulan at 2.6 Å resolution. *J. Mol. Biol.* **287**, 907–921.
- Knegtel, R. M., Wind, R. D., Rozeboom, H. J., Kalk, K. H., Buitelaar, R. M., Dijkhuizen, L. and Dijkstra, B. W. (1996) Crystal structure at 2.3 Å resolution and revised nucleotide sequence of the thermostable cyclodextrin glycosyltransferase from *Thermonaerobacterium thermosulfurigenes* EM1. *J. Mol. Biol.* **256**, 611–622.
- Mikami, B., Adachi, M., Kage, T., Sarikaya, E., Nanmori, T., Shinke, R. and Utsumi, S. (1999) Structure of raw starch-digesting *Bacillus cereus* beta-amylase complexed with maltose. *Biochemistry* **38**, 7050–7061.
- Oyama, T., Kusunoki, M., Kishimoto, Y., Takasaki, Y. and Nitta, Y. (1999) Crystal structure of beta-amylase from *Bacillus cereus* var. *mycoides* at 2.2 Å resolution. *J. Biochem. (Tokyo)* **125**, 1120–1130.
- 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.
- Lawson, C. L., van Montfort, R., Strokopytov, B., Rozeboom, H. J., Kalk, K. H., de Vries, G. E., Penninga, D., Dijkhuizen, L. and Dijkstra, B. W. (1994) Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form. *J. Mol. Biol.* **236**, 590–600.
- 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.
- 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)* **98**, 663–671.
- Bork, P., Dandekar, T., Eisenhaber, F. and Huynen, M. (1998) Characterization of targeting domains by sequence analysis: glycogen-binding domains in protein phosphatases. *J. Mol. Med.* **76**, 77–79.
- Coutinho, P. M. and Henrissat, B. (1999) Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G., Henrissat, B. and Svensson, B., eds), pp. 3–12. The Royal Society of Chemistry, Cambridge.
- 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 Lett.* **272**, 5497–5513.
- Kelley, L. A., MacCallum, R. M. and Sternberg, M. J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**, 499–520.
- Guan, L., Hu, Y. and Kaback, H. R. (2003) Aromatic stacking in the sugar binding site of the lactose permease. *Biochemistry* **42**, 1377–1382.

- 34 Ponyi, 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
- 35 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
- 36 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
- 37 Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. and Barton, G. J. (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**, 892–893
- 38 Combet, C., Blanchet, C., Geourjon, C. and Deleage, G. (2000) NPS@: network protein sequence analysis. *Trends Biochem. Sci.* **25**, 147–150
- 39 Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882
- 40 Peitsch, M. C. (1995) Protein modeling by e-mail. *Bio/Technology* **13**, 658–660
- 41 Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723
- 42 Schwede, T., Kopp, J., Guex, N. and Peitsch, M. C. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385
- 43 Spande, T. F. and Witkop, B. (1967) Determination of the tryptophan content of proteins with *N*-bromosuccinimide. *Methods Enzymol.* **11**, 498–506
- 44 Sokolovsky, M., Riordan, J. F. and Vallee, B. L. (1966) Tetranitromethane. A reagent for the nitration of tyrosyl residues in proteins. *Biochemistry* **5**, 3582–3589
- 45 Pearl, F., Todd, A., Sillitoe, I., Dibley, M., Redfern, O., Lewis, T., Bennett, C., Marsden, R., Grant, A., Lee, D. et al. (2005) The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. *Nucleic Acids Res.* **33**, D247–D251
- 46 Roberge, M., Lewis, R. N., Shareck, F., Morosoli, R., Kluepfel, D., Dupont, C. and McElhaney, R. N. (2003) Differential scanning calorimetric, circular dichroism, and Fourier transform infrared spectroscopic characterization of the thermal unfolding of xylanase A from *Streptomyces lividans*. *Proteins* **50**, 341–354
- 47 Khan, M. Y., Villanueva, G. and Newman, S. A. (1989) On the origin of the positive band in the far-ultraviolet circular dichroic spectrum of fibronectin. *J. Biol. Chem.* **264**, 2139–2142
- 48 Bray, M. R., Carriere, A. D. and Clarke, A. J. (1994) Quantitation of tryptophan and tyrosine residues in proteins by fourth-derivative spectroscopy. *Anal. Biochem.* **221**, 278–284
- 49 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
- 50 Belshaw, N. J. and Williamson, G. (1990) Production and purification of a granular-starch-binding domain of glucoamylase 1 from *Aspergillus niger*. *FEBS Lett.* **269**, 350–353
- 51 Paldi, T., Levy, I. and Shoseyov, O. (2003) Glucoamylase starch-binding domain of *Aspergillus niger* B1: molecular cloning and functional characterization. *Biochem. J.* **372**, 905–910
- 52 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

Received 14 December 2005/8 February 2006; accepted 2 March 2006

Published as BJ Immediate Publication 2 March 2006, doi:10.1042/BJ20051982