Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to β-cyclodextrin

Kay Sorimachi[1], Marie-Françoise Le Gal-Coëffet[1,2], Gary Williamson[2], David B Archer[2] and Michael P Williamson[1]*

Background: Carbohydrate-binding domains are usually small and physically separate from the catalytic domains of hydrolytic enzymes. Glucoamylase 1 (G1) from *Aspergillus niger*, an enzyme used widely in the food and brewing industries, contains a granular starch binding domain (SBD) which is separated from the catalytic domain by a semi-rigid linker. The aim of this study was to determine how the SBD binds to starch, and thereby more generally to throw light on the role of carbohydrate-binding domains in the hydrolysis of insoluble polysaccharides.

Results: The solution structure of the SBD of *A. niger* G1 bound to β-cyclodextrin (βCD), a cyclic starch analogue, shows that the well-defined β-sheet structure seen in the free SBD is maintained in the SBD–βCD complex. The main differences between the free and bound states of the SBD are observed in loop regions, in or near the two starch-binding sites. The two binding sites, each of which binds one molecule of βCD, are structurally different. Binding site 1 is small and accessible, and its structure changes very little upon ligand binding. Site 2 is longer and undergoes a significant structural change on binding. Part of this site comprises a flexible loop, which appears to allow the SBD to bind to starch strands in a range of orientations.

Conclusions: The two starch-binding sites of the SBD probably differ functionally as well as structurally; site 1 probably acts as the initial starch recognition site, whereas site 2 is involved in specific recognition of appropriate regions of starch. The two starch strands are bound at approximately 90° to each other. This may be functionally important, as it may force starch strands apart thus increasing the hydrolyzable surface, or alternatively it may localize the enzyme to noncrystalline (more hydrolyzable) areas of starch. The region of the SBD where the linker to the catalytic domain is attached is flexible, allowing the catalytic site to access a large surface area of the starch granules.

Addresses: [1]Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK and [2]Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK.

*Corresponding author.
E-mail: M.Williamson@Sheffield.ac.uk

Key words: β-cyclodextrin, glucoamylase 1, NMR, solution structure, starch-binding domain

Received: 29 January 1997
Revisions requested: 26 February 1997
Revisions received: 21 March 1997
Accepted: 2 April 1997

Structure 15 May 1997, 5:647–661
http://biomednet.com/elecref/0969212600500647

Introduction

Starch-degrading and related enzymes are abundant in animals, bacteria, fungi and plants [1,2]. The fungal enzyme, glucoamylase 1 (G1; 1,4-α-D-glucan glucohydrolase, E.C. 3.2.1.3) from *Aspergillus niger*, hydrolyses α-D-glucosidic bonds of starch and other polysaccharides to yield β-D-glucose. G1 consists of two functional domains, an N-terminal catalytic domain (residues 1–470, 55 kDa) and a C-terminal granular starch binding domain (SBD; residues 509–616, 12 kDa). The two domains are joined by a bulky, heavily *O*-glycosylated linker (residues 471–508, 13 kDa), composed primarily of serine and threonine residues [3]. Glucoamylases lacking SBDs have unchanged hydrolytic rates against soluble substrates, but have dramatically slower rates against granular starch [4]. Carbohydrate-binding domains with similar characteristics and function (i.e. that are physically separate from the catalytic domain and that bind specifically to insoluble carbohydrate substrates) have been found in various cognate enzymes, for example, chitinase A1 [5], bacterial α and β amylases [6–9], cyclodextrin glycosyltransferase (CGT; [10–11]), xylanase D [12] and glycanases [13].

The presence of these carbohydrate-binding domains in a range of hydrolases clearly highlights their functional importance, although their precise mode of action is not known. Substrate recognition and binding by these domains are extremely specific. For example, xylanase D from *Cellulomonas fimi* has two noncatalytic binding domains that clearly bind to different ligands; xylan binds only to the xylan-binding domain and cellulose is the specific substrate for the cellulose-binding domain. There has been considerable debate over the function of carbohydrate-binding domains. It is clear that a major function is to attach the catalytic domain to its polymeric substrate and thereby increase the effective concentration of substrate at

the active site. It is, however, unclear whether the binding domains also function to prise apart neighbouring carbohydrate strands and thereby make the substrate more accessible for hydrolysis. In order to understand the mechanism by which the substrate specifically interacts with the binding domain and to determine the functional relationship between carbohydrate binding and catalysis, structural information at the atomic level is required.

Only a handful of solution structures of noncatalytic-binding domains have been determined — the type I cellulose-binding domain of cellobiohydrolase I from *Trichoderma reesei* [14], the type II cellulose-binding domain of C_{ex} from *C. fimi* [15] and the type IV cellulose-binding domain of CenC from *C. fimi* [16]. Although all of these have a β-sheet structure, the size of the cellulose-binding domain is not the same between different families and the structure also varies from a three-stranded antiparallel β sheet to a ten-stranded jelly roll β sandwich. Another enzyme of interest is CGT, because, like G1, it also interacts with starch, and its putative granular starch binding domain shows approximately 37% amino acid sequence identity to the G1 SBD. Crystal structures have been determined for free CGT and the CGT–maltose complex from *Bacillus circulans* and *Bacillus stearothermophilus* [11,17,18].

In this paper, we present the solution structure of the SBD from *A. niger* G1 bound to β-cyclodextrin (βCD), a cyclic analogue of starch, and compare it to the solution structure of free SBD [19]. To our knowledge, this is the first solution structure of a carbohydrate-binding domain bound to a ligand. With details obtained from this study, we aim to further our understanding of the nature and specificity of protein–polysaccharide interactions, in particular, the binding of granular starch to the SBD. In addition, we hope to probe the relationship between the catalytic and binding domains in the intact enzyme.

Results and discussion

For the amino acid residue numbering of the SBD, we have adopted that of intact G1 [3]; thus, the SBD comprises residues 509–616. The β-strand residues and numbering in SBD are as for the free protein [19]. For binding site numbering we have followed that of Lawson *et al.* ([11]; Brookhaven Protein Data Bank accession code, 1CDG), which is opposite to the numbering adopted by Klein and Schulz [17] and Coutinho and Reilly [20].

Assignment of the SBD–βCD complex and derivation of constraints

^1H and ^{15}N nuclear magnetic resonance (NMR) assignments were made from multidimensional homonuclear and heteronuclear data as described previously [21]. The assignment process for the SBD–βCD complex was carried out independently to that of free SBD. The difficult regions to assign were mainly those residues in the N terminus (Cys509–Pro512), and those close to the C terminus (Pro601–Gly605). These two groups of residues are close in space due to the presence of a disulphide bond (Cys509–Cys604). The ^1H and ^{15}N assignments of these residues remain tentative or incomplete, primarily due to broadening of resonances as was observed in free SBD [21]. A table of ^1H and ^{15}N chemical shifts is supplied as Supplementary material, which is available from the Internet version of this paper.

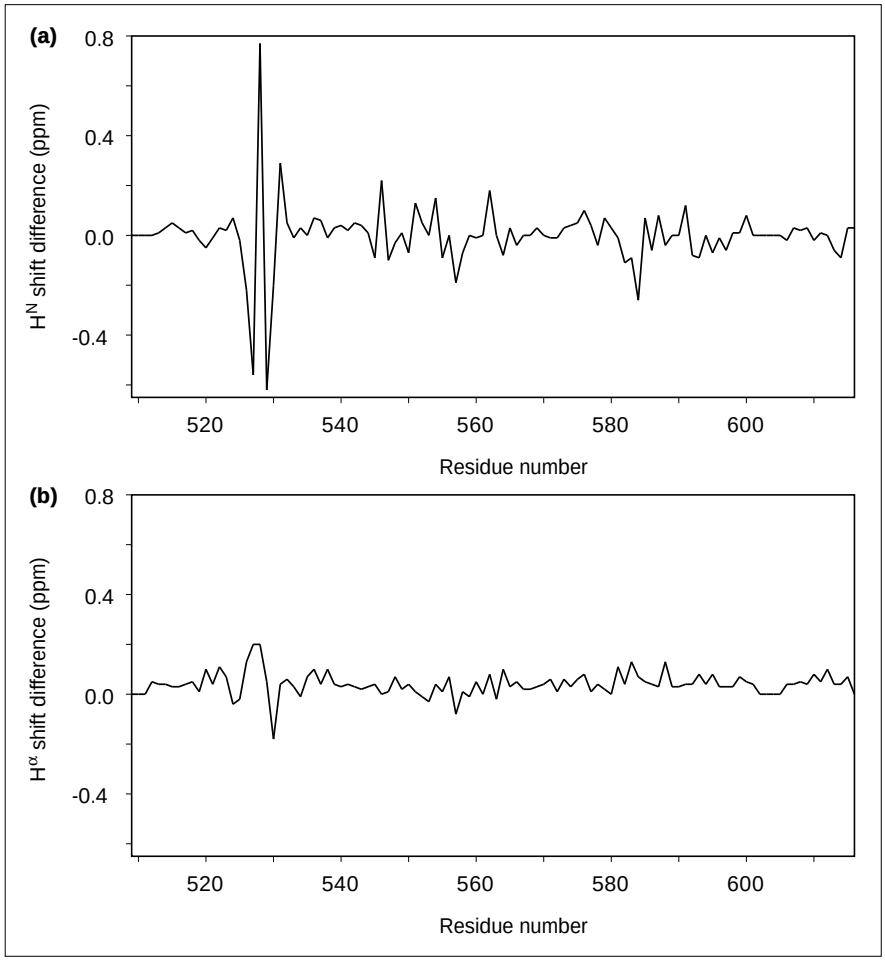
Previously, we reported the observation of two separate spin systems for particular residues in free SBD, which were due to *cis–trans* proline isomerization and/or glycosylation of a nearby residue [21]. In the SBD–βCD complex, an additional residue, Glu573, was also assigned to two spin systems, presumably due to *cis–trans* isomerization of Pro570. Re-examination of spectra showed that the second spin system for Glu573 was probably also present in free SBD, but the crosspeaks are extremely weak. This, however, does not affect the solution structure determination of free SBD [19].

Chemical shifts for βCD were assigned mainly from TOCSY spectra and by comparing one-dimensional (1D) and two-dimensional (2D) spectra of free SBD and the SBD–βCD complex. An assignment table is supplied as Supplementary material, which is available from the Internet version of this paper. Two separate sets of βCD signals were observed, corresponding to the two molecules of βCD bound to the SBD (see later discussions). The structure of βCD has the shape of a truncated cone and is approximately circular in cross-section. This may explain why its proton resonances are in fast exchange on the chemical shift timescale, such that equivalent protons on each of the seven saccharide units have averaged (identical) shifts (although a different shift is observed for each of the two bound βCD molecules. The averaged shifts change by up to 0.36 ppm on binding to SBD, indicating the presence of some large ring-current effects.

Following assignment of the complex, the chemical shifts of backbone H^N and H^α protons of SBD were compared for the free and bound states. The difference in chemical shift between the two states was calculated as (δ_{free}–δ_{bound}). A summary of the calculated difference in chemical shifts is shown in Figure 1. The absolute difference for the H^α protons was, overall, very small with most values falling within 0.1 parts per million (ppm). Slightly larger differences were observed for Thr526 (0.13 ppm), Tyr527 (0.20 ppm), Gly528 (0.20 ppm) and Asn530 (0.18 ppm). The H^N shifts were more sensitive to the addition of ligand. Absolute shift changes greater than 0.1 ppm were observed for 12 residues and, of these, the most notable changes were recorded for Tyr527 (0.42 ppm), Gly528 (0.77 ppm), Glu529 (0.48 ppm), Ile531 (0.29 ppm), Ser546 (0.22 ppm) and Ser584 (0.26 ppm).

Figure 1

A plot of (a) the H^N and (b) the H^α chemical shift difference between free SBD and the SBD–βCD complex. For each residue the shift difference was calculated as (δ_{free}–δ_{bound}) and corrected for the difference in temperature. A value of zero was used where assignment was ambiguous or unavailable (i.e. residues Cys509–Thr511 and Gln602–Gly605, all H^N of proline residues and H^α of Pro561).



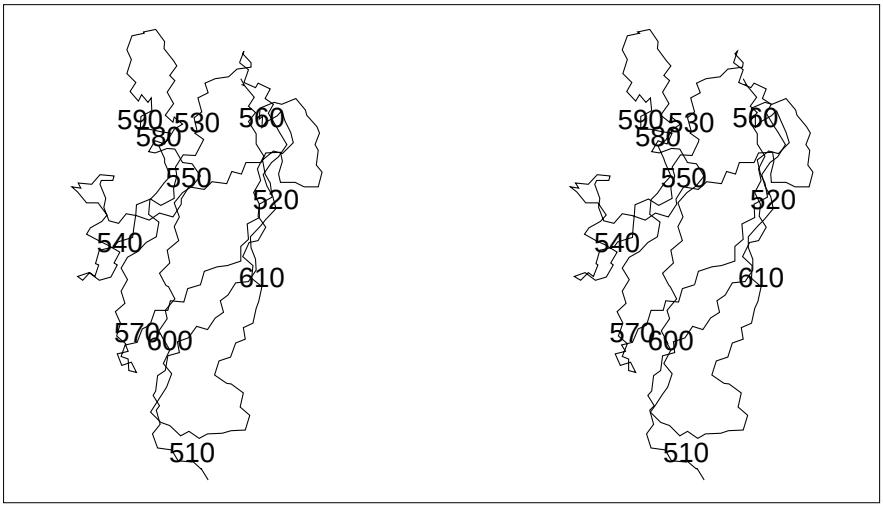
Distance and dihedral-angle constraints were derived as described for the calculation of free SBD [19]. A total of 804 intramolecular, 22 intermolecular and 92 hydrogen bond distance constraints were obtained. Stereospecific assignments were obtained for eight β-methylene pairs and seven valine γ-methyl groups. Dihedral angle constraints were obtained for 58 ϕ and 52 χ_1 angles, giving a total of about 10 constraints per residue.

Binding of βCD to SBD

By NMR and ultraviolet difference spectroscopy the stoichiometry of βCD binding to SBD was shown to be 2:1 (βCD:SBD) [22]. Thus, chemical shift changes were observed on titration of βCD into SBD up to the addition of two mole equivalents of βCD, after which no further changes were seen. This is similar to the homologous domain (E) of CGT, the crystal structure of which shows two distinct sites where the ligand (maltose) binds [11]. We have previously reported the observed ^1H and ^{15}N chemical shift changes of potential SBD binding-site residues during the βCD titration, followed by 1D ^1H and 2D ^1H–^{15}N heteronuclear single quantum correlation (HSQC) spectra [19]. We found that the majority of residues exhibiting large chemical shift changes are located in loops or ends of β strands and can be grouped into two distinct regions corresponding to the two binding sites. Determination of the structure of the SBD–βCD complex now provides us with the opportunity to analyse, in greater detail, the specific interactions between the protein and the carbohydrate ligand.

In binding site 1, observed intermolecular nuclear Overhauser enhancement (NOE) peaks involved sidechain protons of residues Trp543 and Trp590. Both of these have previously been identified as critical binding residues, so this observation comes as no surprise. The majority of the intermolecular NOEs observed in the SBD–βCD complex arise from residues within or around site 2, including Ala523, Thr526, Tyr527, Ala553, Tyr556 and Leu562. Some of the NOEs are weak, in particular those involving residues Ala523 and Thr526, whereas stronger and more extensive interactions are observed to Tyr527.

Figure 2



Stereoview of SBD in SBD–βCD$_{av–min}$. Only the backbone (N, Cα and C) atoms of the protein are shown. Every tenth Cα atom starting from residue 510 has been labelled. The N and C termini are at the bottom and top of the figure, respectively.

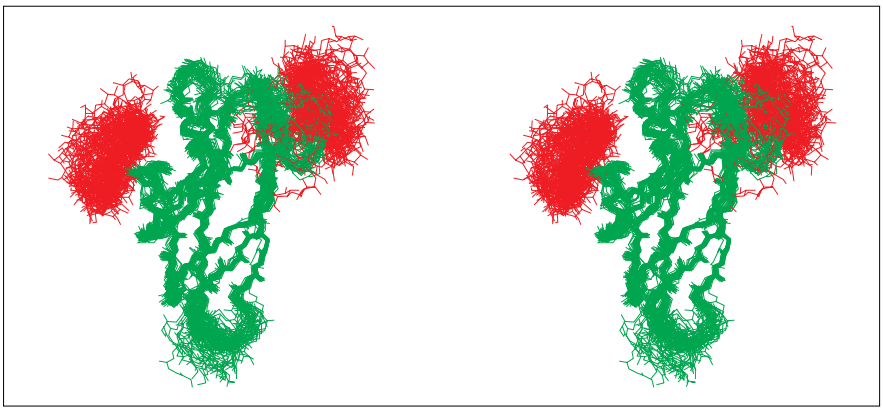
Structure determination

The family of structures for the SBD–βCD complex was calculated using a simulated-annealing protocol. Calculations were performed in an iterative fashion by checking violated constraints and adding new NOE constraints after each stage. Some ambiguous NOE assignments were resolved by inspecting the preliminary low-resolution structures. In total, 22 observed intermolecular NOEs were introduced as distance constraints in the calculation. These constraints were not included at the start, but introduced in the second half of the calculation after the basic fold of the protein had been determined. This resulted in significant saving of computing time, without distorting the final structure of the complex.

In the final calculation, 100 structures were calculated starting from random coordinates and subsequently refined. Resulting structures were ranked in order of total (X-PLOR) energy, NOE energy and root mean square deviation (rmsd) of NOEs. The shape of the curve for all three parameters was almost identical (figure not shown) and similar to those obtained for the free protein [19]. Using the rmsd of NOEs as the selection criterion, the top 5% of structures were rejected immediately due to unreasonably high energies and a large number of unsatisfied constraints. From the remaining structures, 81 had very similar energies and rmsd values. In order to obtain a manageable number of structures for the final analysis, the first (lowest rmsd) and subsequently every second structure were selected, resulting in 41 structures for the family. The average structure (SBD–βCD$_{av}$) was calculated from the 41 final structures and was further subjected to restrained energy minimization to yield the minimized average structure (SBD–βCD$_{av–min}$; Fig. 2).

Figure 3 shows a backbone atom superimposition of the 41 structures to SBD–βCD$_{av}$. The atomic rmsd values for backbone and all heavy atoms when superimposed over

Figure 3



Stereoview of the superimposition of 41 calculated structures of the SBD–βCD complex. The structures were superimposed on to SBD–βCD$_{av}$, on the N, Cα and C atoms of the eight β strands (residues 513–523, 530–536, 549–552, 561–571, 573–582, 589–591, 596–600 and 607–615). The protein (backbone heavy atoms only) and ligand (heavy atoms only) are shown in green and red, respectively. The orientation of the figure is the same as Figure 2.

the eight β-strand residues are 0.54±0.13 Å and 0.92±0.15 Å, respectively. When all residues are used in the superimposition, the atomic rmsd values for backbone and all heavy atoms are 1.19±0.25 Å and 1.60±0.25 Å, respectively. Superimposition of backbone and all heavy atoms of residues with low solvent accessibility (less than 3%; see later discussions) gives rmsd values of 0.45±0.25 Å and 0.82±0.35 Å, respectively. A summary of the structural statistics for the calculated ensemble is given in Table 1.

The number of violations of the NOE constraints (Table 1) is higher than normal for reported NMR structures. This may be explained in two ways: firstly, we have selected 81 from 100 calculated structures to represent the solution structure — this is a higher proportion than is often chosen (where typically the 30% with the lowest violations might be chosen), but we felt that the only logically defensible cutoff is at a point where the convergence of the algorithm deteriorates, rather than after some arbitrarily imposed number of 'best' structures; and secondly, the largest violations involve constraints to βCD. Calculations requiring the docking of two molecules are more difficult, and inclusion of the two βCD molecules led to convergence problems in the simulated annealing calculations. The largest single violation was usually 553Hβ–sugar H3 (with the top two violations for the 41 structures of 1.17 and 0.84 Å). In SBD–βCD_{av-min}, the largest violations are 543δH–sugar H5 (0.57 Å), 553βH–sugar H5 (0.55 Å), and 553βH–sugar H3 (0.53 Å). For the most frequently violated constraints (which all involved βCD), the average violation per structure was in the range 0.35–0.83 Å. There were no significant violations in the more conformationally mobile regions or near to the prolines that display *cis–trans* isomerization.

Table 1

Structural statistics for the SBD–βCD complex of *A. niger* G1.

Parameters	⟨SBD–βCD⟩*	SBD–βCD_{av-min}
NOE violations > 0.5 Å	19.9 ± 3.6	14
Dihedral violations > 5°	6.8 ± 2.3	4
Rmsd from experimental restraints		
distance restraints (Å)	0.141 ± 0.007	0.117
dihedral restraints (°)	2.5 ± 0.4	1.9
Rmsd from idealized geometry		
bonds (Å)	0.0088 ± 0.0009	0.0069
angles (°)	2.56 ± 0.48	5.4
impropers (°)	1.06 ± 0.08	3.1
X-PLOR energies (kcal mol^{-1})†		
E_{NOE}	916 ± 91	625
E_{CDIH}	43 ± 15	23
Atomic rmsd (Å) ⟨SBD–βCD⟩ vs SBD–βCD_{av}	backbone (N, Cα, C)	all heavy atoms
all	1.19 ± 0.25	1.60 ± 0.25
β strands 1–8	0.54 ± 0.13	0.92 ± 0.15
buried residues‡	0.45 ± 0.25	0.82 ± 0.35
SBD–βCD_{av} vs free CGT§		
β strands 1, 2, 4–8	1.91	
β strands 1, 2, 4–7	1.74	
BD–βCD_{av} vs CGT–maltose complex§		
β strands 1, 2, 4–8	1.82	
β strands 1, 2, 4–7	1.66	

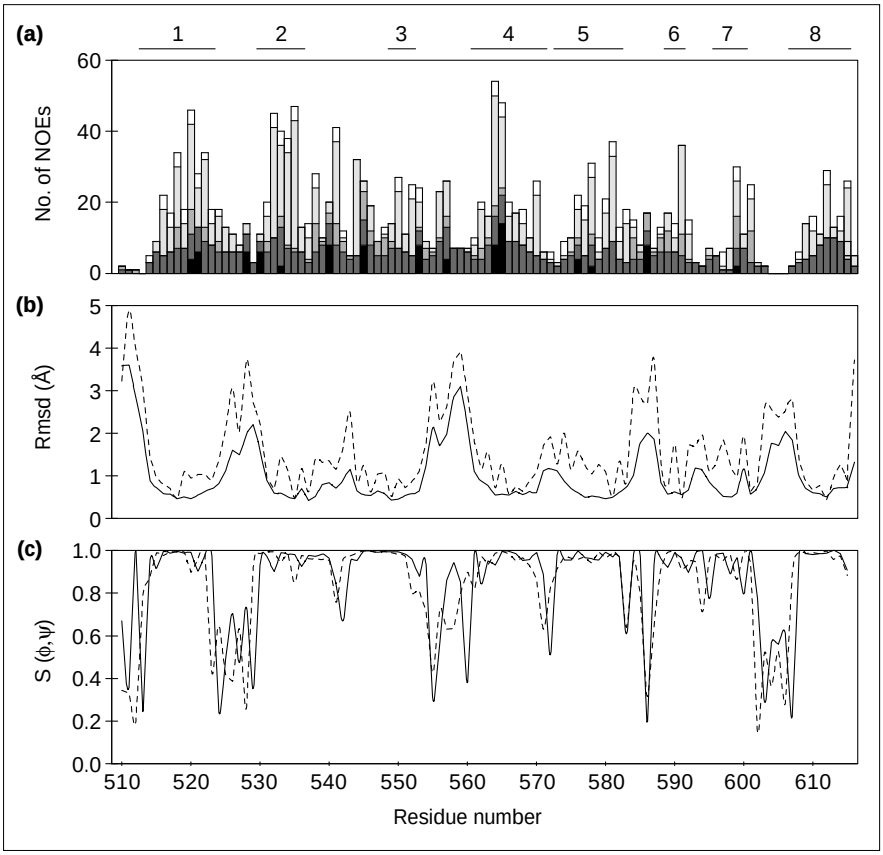
*⟨SBD–βCD⟩, ensemble of 41 final structures. †E_{NOE}, NOE energy term; E_{CDIH}, dihedral angle energy term. Force constants used to calculate E_{NOE} and E_{CDIH} were 50 kcal mol^{-1} Å^{-2} and 200 kcal mol^{-1} rad^{-2}, respectively. ‡Residues with surface accessibility < 3%. §Strand residues common to the SBD–βCD complex, free CGT, and CGT–maltose complex were used in the superimposition.

A graphical summary of the structural statistics is given in Figure 4. The distribution of NOE distance constraints is shown in Figure 4a. A comparative lack of constraints is evident in two regions — the N terminus and the loop spanning residues 601–606. Similar trends were observed for the free protein where fewer constraints were obtained for the N terminus and two loop regions. In Figure 4b, the backbone and sidechain rmsd values for each residue have been plotted. Most of the loop regions clearly have higher rmsd values, as they are not as well-defined as the core of the protein. As in the structure of free SBD, it is evident that the loop between strands 2 and 3 is the best defined loop, despite being the longest loop. Figure 4c is the plot of the angular order parameters (S) of ϕ and ψ dihedral angles. A value approaching unity reflects well-defined local geometry and, therefore, the loops and regions with fewer constraints have smaller values than the more well-defined regions. The 3-residue rmsd values were also calculated (data not shown) and corroborate these trends.

Figure 5 is the Ramachandran plot of ϕ and ψ dihedral angles in the SBD–βCD_{av-min} structure. Although in the SBD–βCD complex there is a greater dispersion of the ϕ and ψ angles and fewer residues in the most favoured regions than in free SBD, there are no residues in disallowed regions.

The surface accessibility was calculated for each residue and those with a value less than 3% were regarded as buried; these are Val515, Val517, Phe519, Leu521, Ile531, Leu533, Val534, Gly535, Leu540, Asp542, Ile549, Leu551, Ser559, Trp563, Leu569, Phe575, Tyr577, Phe579, Ile580, Val600, and Val611. The majority of these residues are hydrophobic in nature as would be expected. The only polar residues in this list are Asp542 and Ser559. The latter has access to solvent through the end of its sidechains (through its Oγ and Hγ atoms). Asp542, although slightly more restricted, can also access solvent. Additionally, its Oδ atoms are close to HN of Lys578, in an orientation that would favour formation of a bifurcated hydrogen bond.

Figure 4



Summary of structural data versus residue number. The β-strand location and numbering are marked at the top. **(a)** The distribution of distance constraints is shown for each residue by bars representing the number of intraresidue (solid bar), sequential, medium-range, long-range NOEs and hydrogen-bond (open bar) constraints, cumulatively stacked with decreasing intensity of shading. **(b)** The average rmsd (Å); backbone, solid line; sidechain, broken line for SBD–βCD_{av}. and **(c)** angular order parameters (ϕ, solid line; ψ, broken line) for the 41 final structures are also presented.

Relaxation experiments

Experiments were carried out to obtain ^{15}N T_1 and T_2 relaxation time constants and NOE values for each residue of SBD in both free and bound states. Values were obtained for 94 residues in free SBD and 93 residues in the complex, and a summary for each residue is shown in Figure 6. The average values for the three parameters, in order of T_1, T_2 and NOE are: 444.6 ms, 118.8 ms and 0.81 for free SBD; 524.7 ms, 99.3 ms and 0.89 for the complex. Using an initial estimate of 6.0 ns for the overall correlation time, the optimized value calculated by the Modelfree program (see Materials and methods) was 7.99 ns and 9.37 ns for free and bound SBD, respectively, which are approximately in line with values expected for monomeric protein.

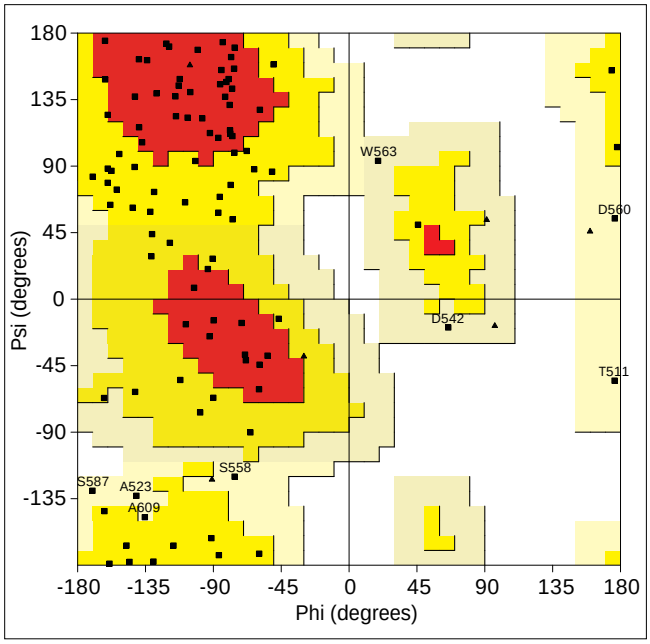
The data show that most of the protein has very limited internal mobility in both free and bound states. In the free SBD solution structures, the least well-defined region spanned residues 509–512 and 601–606 (containing the disulphide bond between Cys509 and Cys604). Many of these residues have ^{15}N resonances that were too broad to permit accurate measurement of relaxation parameters (which in itself is an indication of significant internal motion). Of those resonances that could be measured, Ala603 and Ser606 have relaxation parameters that suggest increased flexibility in this segment. It is significant that Tyr527 (see later discussion) also has ^{15}N resonances too broad to permit measurement of its relaxation parameters.

Solution structure of SBD bound to βCD: global comparison with free SBD

The calculated structures of the SBD–βCD complex show that the protein consists of eight β strands, arranged mainly in an antiparallel fashion. The backbone amide exchange and NOE data are consistent with all residues involved in β strands being the same as those in the free protein [19]. One molecule of βCD is located at each of the two binding sites, as shown in Figure 3.

A comparison of the solution structures of free and bound SBD is shown in Figure 7. The core (β strand) structure of the SBD is essentially the same in the two states, highlighting its contribution to structural stability of this domain. The main structural differences are evident in the loop regions, in particular the loop between strands 2 and 3, as well as that between strands 3 and 4. These loops contain residues involved in ligand binding.

Figure 5

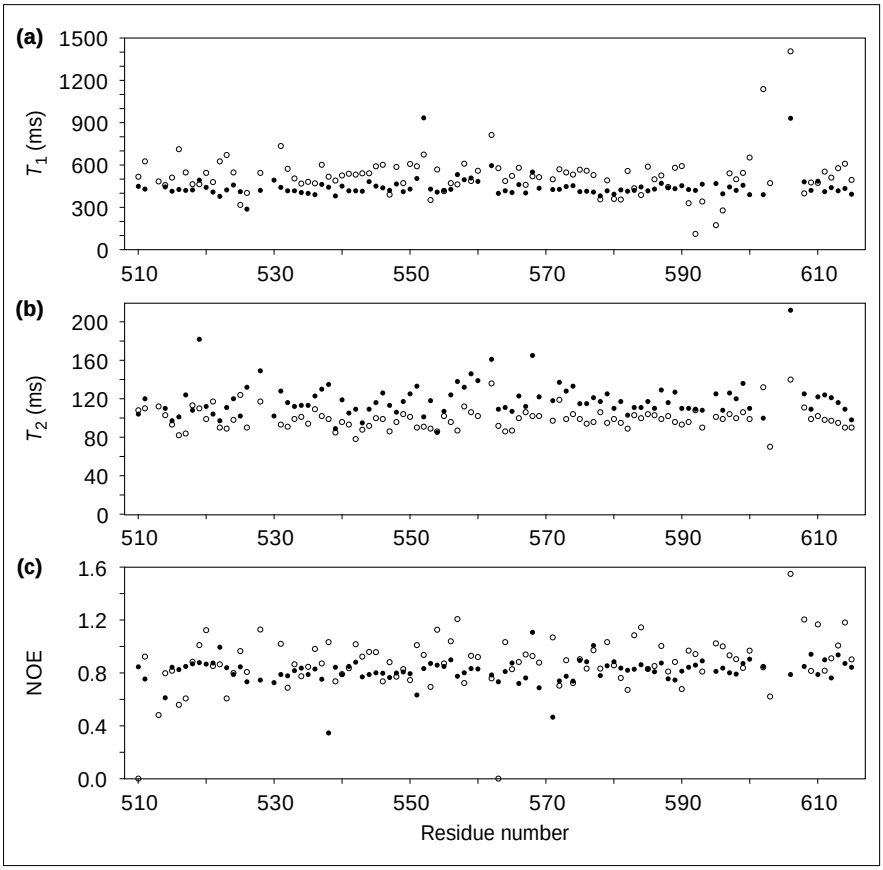


Ramachandran plot of ϕ and ψ dihedral angles for SBD–β$CD_{av–min}$. The plot was generated using the program PROCHECK version 3.4.3 [57]. The coloured areas indicate, from red to light yellow, most favoured region, additional allowed region, generously allowed region, and disallowed region. Residues in the latter two regions are labelled. Glycine residues are shown as triangles, and all other residues are squares.

The backbone amide exchange rates exhibit similar trends in both free and bound SBD (i.e. most residues with fast exchange rates are those not involved in β strands). A qualitative comparison of the rates between the two states shows that the rate is slower for a number of residues in the bound state. The most significant difference is seen for residues Glu591, Arg596 and Asp613, which were classified as medium exchange in free SBD but slow exchange in the complex. These residues, however, are all involved in β strands; this may suggest that the core of SBD is better defined in the presence of the ligand. The loop connecting β strands 2 and 3 overall shows the slowest exchange rates (six residues are classified as medium exchange) of all the loop regions, despite being the longest loop in this domain. In the calculated structures, this region is supported by an extensive network of intraloop hydrogen bonds which help to stabilize the backbone. Another noteworthy residue is Trp563 — its sidechain NH proton (Hε1) shows medium exchange characteristics unlike the other three tryptophan residues. These characteristics were also observed in free SBD although the hydrogen-bonding partner of Trp563 could not be located. In the bound state, a potential hydrogen-bond acceptor has been located as Thr524 N, because in 32 out of the 41 final structures this residue is in a position that satisfies the criteria for hydrogen bonding.

The majority of hydrogen bonds deduced from amide exchange rates and NOE patterns for the complex were similar to those identified in the free protein [19]. These are mostly cross-strand interactions that serve to reinforce the core of the structure. Three additional hydrogen bonds were immediately apparent in the complex — Glu576 HN–Ser536 O, Tyr564 HN–Ser552 O and Phe579 HN–Glu591 O. In each pair, one residue is located at the end of a β strand and, once again, this indicates a better defined, less flexible protein core in the complex. A number of other differences in the hydrogen-bonding pattern were evident following preliminary calculations. Because Ser552 HN is hydrogen bonded to Tyr564 O and shows no NOEs to Ile531 HN, we could not confirm if a hydrogen bond acceptor existed for the slow exchanging Ile531 HN. There is an NOE between Ile531 HN and Leu551 HN, however, and following early calculations, the Ile531 HN–Leu551 O hydrogen bond was confirmed by its presence in all structures. The Glu583 HN–Ser587 O hydrogen bond is also detected in all initial calculated structures and supported by interstrand NOEs. This interaction is located in the hinge region of the loop between β strands 5 and 6 and was also found in free SBD. A number of hydrogen bonds were found in the long-loop region (Ile537–Gly548) between strands 2 and 3 following preliminary calculations — Ile537 HN–Leu540 O, Leu540 HN–Ile537 O and Gly541 HN–Asp547 O. These were all present in 80–100% of the calculated structures, the relevant protons are in medium exchange and there are NOEs supporting these interactions. However, the Leu540 HN–Ile537 O interaction is the only one also present in free SBD and although this long loop is well stabilized in both free and bound states, the pattern of the hydrogen-bond network is not the same in the two states. This probably reflects minor structural changes in this region that accompany ligand binding due to its proximity to Trp543, a key binding residue. From the same loop region, the atom pairs of Asp547 HN–Glu544 O and Gly548 HN–Thr545 O appeared consistently to form hydrogen bonds (according to distance and angle criteria) in all calculated structures. However, these were not supported by H^N-exchange data and were therefore not included in the calculations. Another important tryptophan residue is Trp563, the HN atom of which is in slow exchange and its hydrogen-bond acceptor was deduced to be Leu521 O. Surprisingly, when this hydrogen bond constraint was included in the calculation, it was only satisfied in 10% of the structures and many NOEs supporting this interaction were violated. Instead, the most feasible acceptor was found to be Asp560 O and although this initially looked awkward by eye, other supporting NOEs were found (e.g. Pro561 Hβ–Ala523 HN) and the NOEs found earlier (e.g. Thr522 Hα–Trp563 HN and Ala523 HN–Trp563 HN) were still satisfied. Thus, the Trp563 HN–Leu521 O hydrogen bond constraint was replaced by Trp563 HN–Asp560 O. This new

Figure 6



Experimental ^{15}N relaxation parameters ((a) T_1, (b) T_2 and (c) NOE) obtained for free SBD (closed circles) and the SBD–βCD complex (open circles). Because we only have three experimental parameters recorded at a single magnetic field strength for each protein state, we do not feel that it is justified to present a more detailed analysis of the relaxation data.

hydrogen bond stabilizes the loop between β strands 3 and 4, which contains residues critical to binding site 2. Although not directly involved in ligand binding as Trp543 and Trp590 are in site 1, Trp563 is a critical residue in site 2; it is central to many interactions and stabilizes neighbouring residues which create a surface for the ligand to bind.

In summary, the hydrogen-bonding patterns in free SBD and in the complex are very similar. The main differences are that the β-sheet core is further stabilized in the complex, there is a small structural change of the loop between strands 2 and 3, and the loop between strands 3 and 4 is stabilized.

In the free SBD structure, residue Asp542 displayed unusual backbone geometry (ϕ angle of 81.9°) and was located in the disallowed region of the Ramachandran plot. In the majority of final structures for the complex, Asp542 also has a positive ϕ angle in the range of 63.0°–86.8°. As discussed previously [19], this geometry is thought to be necessary for the binding residue Trp543 to adopt a conformation which would allow interaction with the ligand.

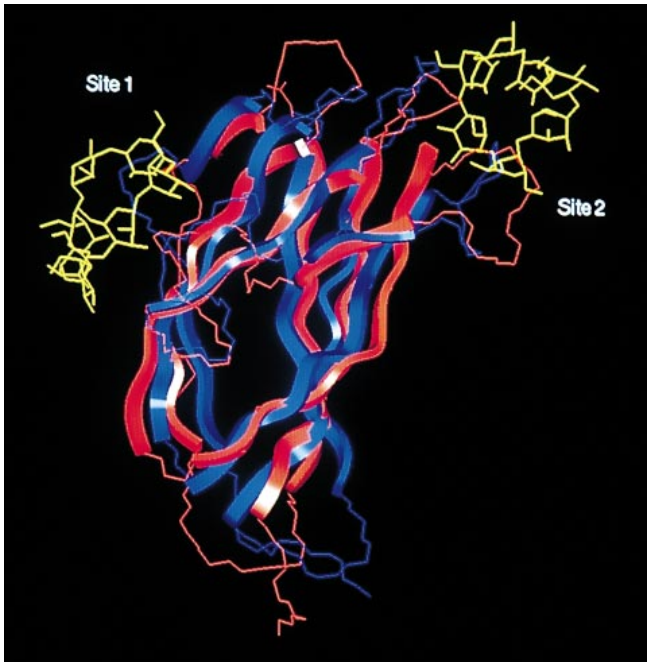
Analysis of binding sites

Cyclodextrin binds to the SBD with the apolar faces of the sugars stacked against aromatic rings — Trp543 and Trp590 in site 1, and Tyr527 and Tyr556 in site 2. Similar observations have been made with CGT [23], α-amylase [24] and β-amylase [25]. Tryptophan residues play an important part in the function of SBD and all four are highly conserved in starch-degrading enzymes. Residues Trp543, Trp563 and Trp590 are all intimately involved in the binding sites, as later discussed. The fourth tryptophan, Trp615, is not involved in either binding site. Nevertheless, it has been shown to be a key residue [26]. We have mutated this tryptophan to a lysine. The resultant variant has greatly reduced stability, implying that Trp615 has a primarily structural role [27].

Binding site 1

Binding site 1 is dominated by Trp543 and Trp590, which form a compact, rigid and surface-exposed hydrophobic site with an inter-ring spacing appropriate for binding to α-1,4-linked glucoses. The structure of this binding site is almost unchanged in the SBD–βCD complex from that in free SBD. It is also closely similar to site 1 in the CGT–maltose complex [11], with respect to both the

Figure 7



A view of the superimposition of SBD–βCD_{av-min} (red) and the minimized average structure of free SBD (blue). The structure of free SBD was superimposed on to the N, Cα and C atoms of β strands 1–8 of SBD–βCD_{av-min}. The βCD molecules are shown in yellow.

positions of key tryptophans and the orientation of the binding site relative to the protein core. The cyclic nature of the ligand and the ambiguous definition of the intermolecular NOEs, which allows for the possibility that any of the sugars may interact with the protein (see Materials and methods section), make the flat stacking of the sugar ring onto the hydrophobic tryptophan indole ring less parallel than expected. Nevertheless, both Trp543 and Trp590 are in close proximity to the sugar rings as the NOEs indicated and the van der Waals surface of these protein residues and the ligand are closely packed together. The saccharide units that stack against the two tryptophans are units *i* and *i*+2 of the ligand. Potential hydrogen bonds are also apparent which would strengthen this interaction. The structure of starch is not known in detail (though models have been proposed; see later discussions), so it is unwise to propose a detailed hydrogen-bonding scheme for the SBD–starch complex on the basis of the structure of the SBD–βCD complex. However, it is likely that all of the residues identified as potential βCD ligands can form hydrogen bonds with starch, although some interactions may be mediated by water molecules.

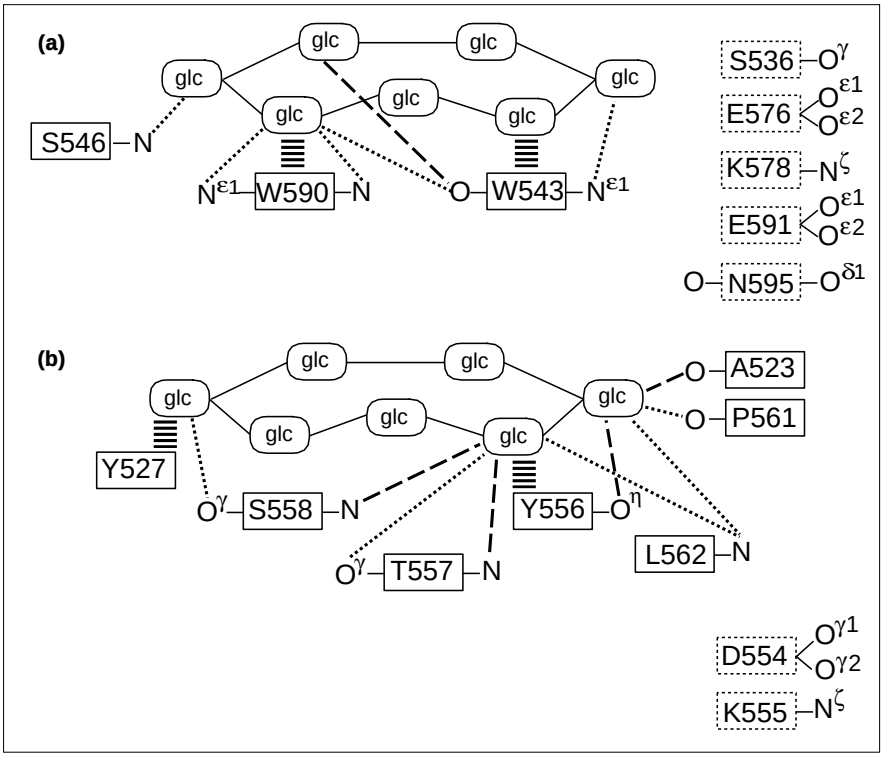
We previously proposed potential binding residues on the basis of βCD titration studies and from modelling the free SBD structure onto the crystal structure of the CGT–maltose complex [19]. The two tryptophan residues discussed above are in similar positions in the free and bound SBD structures. The three other residues previously proposed, Lys578, Glu591 and Asn595, all occupy different conformational space in the complex than in free SBD, but we could not assign constraints for them beyond Hβ. Thus, although these residues do not form the same interactions in the current structure as in the previously proposed model, our data do not define the locations of the sidechains. As the sidechains of these three residues protrude outwards forming part of the protein surface, they would still be capable of interacting with a large ligand such as starch. Figure 8a gives a summary of the possible intermolecular interactions in binding site 1.

Binding site 2

Binding site 2 is more extended than site 1 and undergoes conformational rearrangement on binding to βCD, particularly in the loop between strands 3 and 4 and at Tyr527, which forms the apex of the loop between strands 1 and 2. The protein–ligand interactions in this site are once again dominated by hydrophobic effects from two aromatic rings, Tyr527 and Tyr556. These two planar rings pack closely against the van der Waals surface of the βCD molecule. An inspection of all 41 final structures shows that the position of the ligand is more variable in this binding site; this appears to be due to the flexibility of residue Tyr527. Although Tyr556 also interacts closely with βCD, the position of this tyrosine in the ensemble of structures is reasonably static. Residue Tyr527, on the other hand, appears to accommodate movement in the ligand that would occur in solution. The flexibility of this residue was already evident in the free protein, where two ring conformations were observed and therefore the stacking interaction with the ligand, when modelled onto the CGT–maltose complex, was not well defined [19]. It is possible that with a larger ligand/substrate, the flexibility observed in Tyr527 would be reduced, as there would be a larger protein–ligand interaction surface. It is additionally feasible that the flexibility associated with Tyr527 is functional, in that it can act to locate the ligand and guide it towards the binding site.

Residue Tyr556 in the free SBD structure is in a very similar position to a tyrosine residue in the CGT structures (Tyr632 in free CGT; Tyr633 in the CGT–maltose complex), although these are not the corresponding homologues of Tyr556 in CGT. In the SBD–βCD complex, Tyr556 has moved closer to Asp554 and Lys555 (corresponding to Asn626 and Gln627, respectively, in free CGT) and its aromatic ring would be located near the sidechains of these CGT residues. These residues (and their homologues Asn627 and Gln628, respectively, in the CGT–maltose complex) have been identified as binding-site residues in free [17] and bound [11] CGT. It appears that this shift in location of Tyr556 in the SBD complex is

Figure 8



Schematic diagram representing SBD–βCD intermolecular interactions suggested in **(a)** binding site 1 and **(b)** site 2. The βCD molecule is shown by seven units marked 'glc'. Hydrophobic-ring stacking is represented by thick horizontal lines between the sugar and the aromatic ring. Direct hydrogen bonds are shown as dashed lines; dotted lines represent possible water-mediated hydrogen bonds. Residues enclosed in dotted square boxes are those at the surface of the protein which may interact with a larger ligand (see text).

related to the position of Asp560, which has changed substantially (the Cα atom has moved more than 13 Å from its position in free SBD). Despite this large conformational change, the indole ring of Trp563 is maintained in approximately the same position as in free SBD and the CGT structures. Thus, Asp560 may be contributing to the stabilization of this region by hydrogen bonding to Trp563 and thereby positioning the loop containing Tyr556 to interact with the ligand.

Although Trp563 is classified as a buried residue (it has negligible surface accessibility) and is incapable of a direct interaction with βCD, it is in close contact with many residues in and around binding site 2. There is a slight difference in the location of the indole ring of Trp563 between the free and bound states of the SBD. This is not a large change and many features observed in free SBD are retained. For example, Ile531 HN is still pointing towards the sidechain of Trp563 and this is reflected in the upfield-shifted chemical shift (5.12 ppm), compared to a similarly upfield-shifted 5.48 ppm in the free SBD. The main residues affected by the slightly different tilt angle of the indole ring are Thr526–Ile531; the indole ring rests against the backbone in this region. The change in backbone conformation is further highlighted by the large chemical shift changes of Tyr527 on ligand binding.

In binding site 2, there are a number of residues on the surface of the protein that are within direct or water-mediated hydrogen-bonding distance of a potential starch ligand. In addition, Asp554 and Lys555 have their sidechains protruding outwards, such that they could potentially form an interaction with a larger ligand. Some residues (e.g. Thr525, Thr526) that were flexible on the surface in free SBD, are packed tighter together upon ligand binding; this packing contributes to the stability of the binding interface. The residues in site 2 form a larger surface area for the ligand to interact with than those in site 1. This may be related to the different roles/functions of the two sites. A summary of the possible intermolecular interactions observed in binding site 2 is shown in Figure 8b.

Comparison with crystal structures of other carbohydrate-binding proteins

In addition to comparing the structure of the SBD–βCD complex to the free SBD solution structure, we also made comparisons to a number of crystal structures of other carbohydrate-binding proteins. As noted previously [19], the global fold of the putative starch-binding domain of CGT is very similar to that of SBD. The main difference is that the third β strand of SBD is not evident in CGT. Both proteins possess two binding sites, and many key binding residues are conserved. The four tryptophan residues are found in similar positions for both proteins in their free

and bound states. In binding site 1, the position of Glu591 is also very similar to the position of its homologous binding residue Glu663 in the CGT–maltose complex. Inspection of Glu576 in the SBD–βCD complex shows that this charged residue also has its sidechain at the surface close to binding site 1 and it may be capable of forming an interaction with the ligand–substrate. The observed chemical shift changes upon addition of βCD for Glu576 HN and Hα are 0.10 ppm and 0.08 ppm, respectively, which are moderate changes. The homologous residue in free CGT (Glu648) was in fact identified as a binding residue [17].

The principal protein–ligand interaction seen in both binding sites is face-to-face stacking of aromatic rings against the hydrophobic sugar face. Aromatic residues dominating the interaction between ligand and protein have also been seen in the structure of the Fab fragment of the Se155-4 antibody bound to an oligosaccharide [28]. Three sugars from βCD are involved in the interaction with SBD at binding site 1 and four interact at site *2*, although only two sugars in each site make direct contact with aromatic residues of the protein. The number of direct protein–ligand contacts is perhaps fewer than expected, although it is possible that some intermolecular hydrogen bonds mediated by water molecules exist. It is interesting to note that in the crystal structure of the maltodextrin-binding protein–βCD complex [29], the ligand made 94 interactions of less than 4.0 Å with the protein and bound water molecules. However, there were only four direct sugar–protein hydrogen bonds plus another four water-mediated hydrogen bonds. The number of βCD sugar residues involved in protein–ligand interactions in this complex was three. Both this maltodextrin-binding protein and the SBD highlight the largely nonspecific interactions encountered in protein–βCD complexes which, in fact, appears to be the case for many protein–carbohydrate structures.

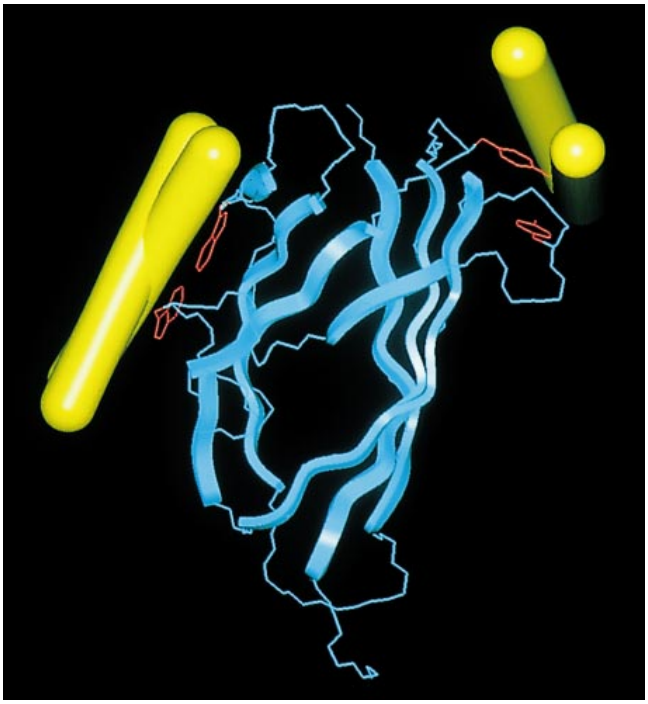
In a recent crystallographic study of CGT from *B. circulans* strain 251 bound to α-cyclodextrin, residue Leu600 was found to be approximately in the centre of the ligand molecule [23]. In CGT from *Thermoanaerobacterium thermosulfurigenes* EM1, this residue has been replaced by Tyr597 [30]. The homologous residue in SBD is in fact Tyr527, which we have observed to be closely associated with βCD and plays a major role in binding site *2* in the structure of the SBD–βCD complex.

Understanding the binding of starch to SBD

The ultimate aim of this study is to have a better understanding of how granular starch interacts with the SBD. It is clear that the two binding sites in G1 SBD are structurally different. Site 1 is small and rigid, with Trp543 and Trp590 forming a well-exposed planar aromatic surface that is well placed to interact rapidly with carbohydrate, with minimal structural rearrangement. Similar planar aromatic surfaces have been observed in cellulose-binding domains [31]. By contrast, site *2* is longer and undergoes a change in conformation on binding. It is striking that the key aromatic residues in this site, Tyr527 and Tyr556, do not lie parallel to the protein surface (as do the aromatic rings in site 1) but project out from it, as if to lock the ligand into place. Thus, the structure of the complex and the different flexibility of the two binding sites suggest different functional roles for the two sites. A similar observation has also been reported for the two sites in CGT, in which the binding site containing Trp616 and Trp662 (corresponding to the SBD residues Trp543 and Trp590, respectively) is proposed to be largely responsible for attaching the enzyme to raw starch granules, whereas the second binding site located near Tyr633 (close to Trp563 in SBD) may be involved in guiding the starch chain towards the active site [30,32]. Information obtained from our calculated ensemble of structures fully supports such a result, and suggests a model in which binding site 1 acts as the initial starch-recognition region, allowing the SBD to scan rapidly through many potential binding sites until site *2* (and the highly mobile Tyr527 in particular) adopts the appropriate conformation to bind tightly and cooperatively to a second starch strand. This model is reinforced by the binding constants of βCD to the two sites [27], which show that site 1 binds βCD approximately four-fold less strongly than site *2*.

Using the information we have gained from the structure determination of the SBD–βCD complex, a model of an SBD–starch complex was built. Figure 9 shows this model with the protein coordinates retained from SBD–βCD_{av-min} and starch molecules replacing the βCD ligand; it highlights the different orientations that the starch helix can adopt. The interaction in site 1 is compact and quite rigid as shown above for SBD–βCD. The major interactions between SBD and starch only involve two glucose units, which stack over Trp543 and Trp590. Site *2* displays a more extended, broader area of interaction surface. The starch molecules in this site span a much larger conformational space. The main protein–carbohydrate contacts involve Tyr556, which forms a fixed pivot, and Tyr527, which can swing around in an arc, taking the starch strand with it. The starch helices in the two binding sites adopt a markedly nonparallel geometry; the difference in orientation is approximately 90°. This is significant, because it is thought that the starch helices on the surface of starch granules are normally parallel; certainly this must be true for the crystalline regions of starch. Two quite different functional reasons for the nonparallel geometry are plausible — either the SBD is forcing two starch strands apart and thus assisting in the breakdown of starch granules; or the SBD localizes the enzyme to noncrystalline (and therefore presumably more easily hydrolyzable) regions of starch. The former reason has been postulated as a

Figure 9



A view of the SBD–starch model. Two representative orientations of the starch helix are shown in yellow. The aromatic contact residues from SBD–βCD_{av-min} are highlighted in red.

function for the cellulose-binding domain in *C. fimi* endoglucanase A [33], but it is unclear whether this is a general mechanism. We are currently investigating which, if either, of these two mechanisms occurs in G1.

The results presented here show that the 'bottom end' of SBD, as shown in Figures 2 and 3 (i.e. the N terminus and the adjacent loop between strands 7 and 8, containing residues 601–606), is more mobile than the rest of the protein. It is significant that this is the opposite end of the SBD from the end that binds starch, and it is the point of attachment with the catalytic domain. This would imply that the SBD can bind tightly and rigidly to starch, but it can still 'waggle its bottom' and allow the catalytic domain to scour a large area on the surface of the starch granule, thereby combining tight and specific binding with access of the catalytic site to a large number of potential substrates.

Biological implications

Starch is the most abundant storage polysaccharide in plants; it consists of glucose units mainly in the form of amylose and amylopectin, which are thought to be arranged in semicrystalline arrays of double-helical strands to form large irregular granules. The detailed structure of the granules probably differs depending on the producing organism. Starch-degrading enzymes are abundant in animals, plants and microorganisms. Many processes in the food and beverage industries that produce alcohol, soy sauce, syrups and sweeteners rely heavily on fungal enzymes that degrade starch to sugars. Glucoamylase from *Aspergillus niger* is one such enzyme that is used widely in these processes.

Glucoamylase has a structure typical of many polysaccharide-degrading enzymes, with a catalytic domain and a separate polysaccharide-binding domain. The binding domain attaches the enzyme to its polymeric substrate, thereby increasing the concentration of substrate at the active site and dramatically increasing the rate of reaction. It is not clear whether binding domains have other functions; for example they may disrupt the saccharide surface or target the enzyme to particular sites. Such roles have been proposed for some cellulases and hemicellulases.

Our structure determination of the complex between the glucoamylase starch-binding domain (SBD) and **β**-cyclodextrin provides us with a high-resolution solution structure of a carbohydrate-binding domain bound to a ligand. There are two independent binding sites in the SBD that are structurally and probably functionally different. Site 1 is smaller, more rigid and more surface-exposed than site 2, and it is proposed to act as the initial recognition site for starch, whereas site 2 has a larger surface area, undergoes a conformational change on binding, and is proposed to act as the more specific site which acts to lock the SBD into place. Site 2 is also capable of recognizing a range of orientations of starch strands. The two starch strands are bound to the SBD in an unexpected perpendicular orientation, implying that the SBD does more than merely locate the catalytic domain onto the starch surface. We have also shown that although the SBD binds tightly to starch, its site of attachment to the linker, connecting the SBD to the catalytic domain, is flexible, thereby allowing the catalytic domain to access large areas of the starch granule surface. The SBD thus has several simultaneous functions, which may also be relevant to other carbohydrate-binding domains.

Materials and methods

Sample preparation

Unlabelled SBD was obtained by proteolytic cleavage of G1 (Sigma) followed by purification, and uniformly ^{15}N-labelled SBD was expressed in *A. niger* with a pIGF fusion vector, using $^{15}NH_4Cl$ as the sole nitrogen source. These procedures have been described in detail elsewhere [22,34]. The SBD–βCD complex was prepared by addition of the appropriate amount of βCD to achieve a ratio of 1:2 (SBD:βCD). Typically, samples for NMR experiments contained the complex (1–2 mM unlabelled SBD or 1 mM ^{15}N-labelled SBD) in 90% H_2O/10% D_2O or 99.9% D_2O solution.

NMR experiments and data processing

Spectra were recorded on a Bruker AMX 500 or 600 spectrometer at 310K. Additional 2D spectra were recorded at 300K and 320K to resolve ambiguities arising from overlap. The H_2O signal was suppressed

by low-power presaturation during the recycle delay and the ^{1}H carrier was placed at the solvent frequency. ^{1}H chemical shifts were referenced to the H_2O resonance at 4.56 ppm (310K) relative to sodium 3-trimethylsilyl-2,2,3,3-($^2H_4$)propionate. ^{15}N chemical shifts were referenced indirectly by using the ^{1}H frequency for the H_2O resonance and the gyromagnetic ratios [35].

The following multidimensional experiments on the SBD–βCD complex samples were recorded and processed as described previously for free SBD [19,21], except where noted — double quantum filtered correlation spectroscopy (DQF-COSY; [36]), nuclear Overhauser enhancement spectroscopy (NOESY; [37,38]), clean total correlation spectroscopy (TOCSY; [39]), double quantum spectroscopy [40], primitive exclusive correlation spectroscopy (P.E.COSY; [41]), HSQC [42], 3D NOESY-heteronuclear multiple quantum correlation (HMQC) experiment [43], and 3D TOCSY-HMQC experiment [44]. Homonuclear 2D experiments were recorded with 474–750 real t_1 increments. The P.E.COSY was recorded with 500 (real) t_1 increments of 2048 (complex) data points, using the ^{15}N-labelled sample that had been exchanged in D_2O. The NOESY-HMQC was recorded with 220 (complex, t_1), 64 (real, t_2) and 256 (complex, t_3) points. Mixing times of 50–80 ms were used for the acquisition of clean TOCSY spectra and 60–200 ms for NOESY spectra. For the backbone amide exchange experiments, a total of seven ^{15}N HSQC spectra were recorded at 310K, with 64 scans, 256 t_1 increments and 2048 complex points. The time-domain data of homonuclear 2D experiments were apodized by a 60°-shifted sine-squared function (except for the P.E.COSY where a 30°-shifted sine-squared function was used), whereas the 3D data were processed with a 60°-shifted sine-bell function in all dimensions. The first point in each 3D experiment was divided by two before Fourier transformation to suppress ridges in the transformed spectra [45].

Sensitivity-enhanced heteronuclear experiments described by Kördel *et al.* [46] were acquired to measure ^{15}N T_1 and T_2 relaxation time constants and {^{1}H}–^{15}N NOEs. T_1 and T_2 experiments were recorded using eight (0.04, 0.12, 0.2, 0.4, 0.8, 1.2, 2.0 and 3.0 s) and seven (0.004, 0.032, 0.128, 0.22, 0.36, 0.8 and 1.3 s) delays, respectively, with a minimum of two time delays duplicated in each set. All three experiments (except where specified) were recorded with the following parameters: spectral width of 6250 Hz with the carrier placed at the centre of the spectrum, 480–600 increments in t_1, eight (T_1 and T_2) or 32 (NOE) transients per increment, and a recycle delay of 3 s (T_1 and T_2) or 4.4 s (NOE) to ensure sufficient recovery of the ^{1}H magnetization. The dynamics data were analysed by applying a non-linear least-squares fit to single exponential functions with the Modelfree program (version 3.1; [47,48]), which uses a Levenburg–Marquardt algorithm. An optimized value for the overall molecular correlation time (τ_m) was obtained for free and bound SBD. An initial estimate of 6.0 ns was used based on published correlation times of proteins of similar size [15,49].

NMR spectra were generally assigned manually using 2D homonuclear data, initially in conjunction with the software programme PRONTO/3D [50]. The 2D and 3D heteronuclear data were used to confirm the 2D assignments and to help resolve ambiguities arising from overlap of the backbone amide resonances.

The degree of exchange of backbone amide protons was measured using the ^{15}N-labelled protein sample. The sample in H_2O was lyophilized, redissolved in D_2O at time = 0 and a series of 2D ^{1}H–^{15}N HSQC spectra was recorded over a period of five days. The midpoint of each experiment was at the following times: 3.3, 8.6, 11.9, 16.2, 20.5, 44.9 and 119.1 h. The sample was stored in a water bath at 310K between experiments. Those residues with H^N resonances observed in the last spectrum (midpoint time of 119.1 h) and those which were observed in the first spectrum (midpoint time of 3.3 h) but not in the last were deemed to be involved in slow and medium exchange, respectively.

Structure calculations

Distance and dihedral angle constraints, stereospecific assignments, hydrogen bond constraints and the structure calculation protocol used in this study were similar to that used for the structure calculation of free SBD. These were reported in detail previously [19], so only a brief description is given here.

Distance constraints were derived manually from 2D NOESY spectra acquired with mixing times of 60 ms and 200 ms and calibrated using NOEs typical of regular secondary structure. The calibrated upper bounds used for the following number of crosspeak contour levels were: ≥6 contour levels, 2.7 Å; 5 levels, 3.1 Å; 4 levels, 3.5 Å; 3 levels, 4.1 Å; 2 levels, 4.7 Å; 1 level, 5.8 Å. Lower bounds were set to 1.9 Å in all cases. The final set of NOE distance constraints consisted of 45 intraresidue (i = j), 282 sequential (|i–j| = 1), 76 medium-range (2 ≤ |i–j| ≤ 4), 401 long-range (|i–j| ≥ 5) and 22 intermolecular NOEs.

Coupling constants between backbone H^N and H^α ($^3J_{N\alpha}$) were measured from DQF-COSY and NOESY spectra using PRONTO/3D or using the line-fitting interface in the FELIX software package (Felix User Guide, version 2.3, Biosym Technologies, San Diego, 1993). In the structure calculations, for residues with $^3J_{N\alpha}$ values ≥8.0 Hz, <8.0 Hz but ≥7.0 Hz, and ≤5.5 Hz, the corresponding ϕ angles were constrained to $-120° \pm 40°$, $-120° \pm 50°$ and $-65° \pm 25°$, respectively. Stereospecific assignments of prochiral β-methylene groups and valine γ-methyl groups were obtained wherever possible. Restraints for χ_1 and χ_2 angles were set to ± 30° or 40° of the appropriate staggered (180°, –60°, 60°) rotamer as described previously [19].

Hydrogen-bond constraints were not included in initial rounds of calculations. Following preliminary calculations possible hydrogen-bond donor–acceptor pairs were identified and compared to amide exchange and NOE data. In subsequent calculations, these were included as distance constraints of d(H–O) = 1.8–2.5 Å and d(N–O) = 2.5–3.3 Å (i.e. two constraints per hydrogen bond) together with other previously unidentified long-range pairs that were present in 70–100% of the structures and were consistent with amide exchange and NOE data.

Structure calculations were performed with a simulated annealing protocol using the X-PLOR program [51], as previously described for free SBD [19]. The protocol was modified slightly from the free SBD calculation, in that the *dgsa.inp* and *refine.inp* steps were run initially with only the protein–protein NOEs, then repeated with the incorporation of protein–ligand NOEs. Dihedral angle constraints were included from the start. The initial annealing temperature, the number of dynamics steps at high temperature and the number of cooling steps generally had to be increased from the values used in the free SBD calculation. The σ value for oxygen type O was increased from 2.7755 to 3.0068 to reduce the number of bad non-bonded contacts [52]. The sum averaging option was used to treat equivalent and nonstereospecifically assigned protons, to maintain consistency with calculations on free SBD, and because this method (or the essentially equivalent r^{-6} averaging method) is generally the most appropriate in such cases [53]. Due to the cyclic nature of the ligand molecule and the fact that rotation of βCD about its symmetry axis leads to averaging of shifts from the seven saccharide units, intermolecular NOEs could not be assigned to one specific proton of a particular sugar residue. Thus, these were defined between one SBD proton (or a group of equivalent protons) and a group of βCD protons (i.e. the same atom from each of the seven sugars), resulting in a constraint format similar to one which would be used for multiple or ambiguous assignments [54]. For NOEs involving H5 or H6 sugar protons, 21 protons had to be defined to cover all possibilities due to their similar chemical shifts.

Analysis of structures

The calculated structures were viewed and analysed on a Silicon Graphics workstation using the Insight II® molecular modelling system. From the final 41 refined structures, SBD–βCD_{av} was calculated using X-PLOR; this was then energy-minimized by 1500 steps of restrained Powell minimization to obtain SBD–βCD_{av-min}. Individual structures from the ensemble were superimposed onto SBD–βCD_{av}.

Dihedral angle order parameters (S) for ϕ and ψ angles were calculated according to Hyberts *et al.* [55]. A well-defined dihedral angle is reflected by an S value approaching unity. The solvent accessible area for SBD–βCD_{av} was calculated in X-PLOR, using a probe radius of 1.6 Å. For each residue, the accessibility was expressed as a percentage of its total area.

Building the model of SBD–starch complex

A model of an SBD–starch complex was built using the SBD–βCD_{av-min} structure, two calculated structures selected from the final ensemble, and model parameters of the B-starch double helix proposed by Imberty and Perez [56]. The two structures were chosen by eye, following superimposition of the complete ensemble onto SBD–βCD_{av-min}, as those representing the conformational space that the ligand (βCD) can occupy. The structures were superimposed onto the β-strand residue backbone atoms of SBD–βCD_{av-min}. Two model starch molecules were then superimposed onto βCD in each binding site for both structures (i.e. four starch molecules). This was achieved by overlaying two glucose units from one face of the starch helix onto the βCD glucose units in each binding site that are involved in a stacking interaction with SBD aromatic residues (see Results and discussion section). No energy minimization was carried out during this protocol. The starch molecules were then represented simply as cylinders running parallel to the helical axes. Thus in Figure 9, this final SBD–starch model is represented by the SBD structure from SBD–βCD_{av-min} and four starch molecules (two in each binding site).

Accession numbers

The coordinates of five final structures selected at random, coordinates of SBD–βCD_{av-min} and the list of NMR constraints have been deposited in the Brookhaven Protein Data Bank, with the codes 1acz and 1ac0, respectively.

Supplementary material

Supplementary material contains a table of the ^1H and ^{15}N resonance assignments for the SBD–βCD complex and a table of ^1H resonance assignments for βCD in complex with SBD and free βCD.

Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC; LINK grant no. LR50/587 and grant no. 50/C673-1) and by the EU (project no. BIO2-CT94-3008). We thank Jeremy Craven for the use of his data-fitting and plotting programs.

References

1. Svensson, B. (1988). Regional distant sequence homology between amylases, α-glucosidases and transglucanosylases. *FEBS Lett.* **230**, 72–76.
2. Jespersen, H.M., MacGregor, E.A., Sierks, M.R. & Svensson, B. (1991). Comparison of the domain-level organization of starch hydrolases and related enzymes. *Biochem. J.* **280**, 51–55.
3. Svensson, B., Larsen, K., Svendsen, I. & Boel, E. (1983). The complete amino acid sequence of the glycoprotein, glucoamylase G1, from *Aspergillus niger*. *Carlsberg Res. Commun.* **48**, 529–544.
4. Takahashi, T., Kato, K., Ikegami, Y. & Irie, M. (1985). Different behavior towards raw starch of three forms of glucoamylase from a *Rhizopus* Sp. *J. Biochem.* **98**, 663–671.
5. Watanabe, T., Ito, Y., Yamada, T., Hashimoto, M., Sekine, S. & Tanaka, H. (1994). The roles of the C-terminal domain and type III domains of chitinase A1 from *Bacillus circulans* WL-12 in chitin degradation. *J. Bacteriol.* **176**, 4465–4472.
6. Itkor, P., Tsukagoshi, N. & Udaka, S. (1990). Nucleotide sequence of the raw-starch-digesting amylase gene from *Bacillus sp.* B1018 and its strong homology to the cyclodextrin glucanotransferase genes. *Biochem. Biophys. Res. Commun.* **166**, 630–636.
7. Bahl, H., *et al.*, & Antranikian, G. (1991). α-Amylase of *Clostridium thermosulfurogenes* EM1: nucleotide sequence of the gene, processing of the enzyme, and comparison to other α-amylases. *Appl. Environ. Microbiol.* **57**, 1554–1559.
8. Nanmori, T., Shinke, R., Aoki, K. & Nishira, H. (1983). Purification and characterization of β-amylase from *Bacillus cereus* BQ10-S1 Spo II. *Agric. Biol. Chem.* **47**, 941–947.
9. Kitamoto, N., Yamagata, H., Kato, T., Tsukagoshi, N. & Udaka, S. (1988). Cloning and sequencing of the gene encoding thermophilic β-amylase of *Clostridium thermosulfurogenes*. *J. Bacteriol.* **170**, 5848–5854.
10. Nitschke, L., Heeger, K., Bender, H. & Schulz, G.E. (1990). Molecular cloning, nucleotide sequence and expression in *Escherichia coli* of the β-cyclodextrin glycosyltransferase gene from *Bacillus circulans* strain no. 8. *Appl. Microbiol. Biotechnol.* **33**, 542–546.
11. Lawson, C.L., *et al.*, & 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.
12. Millward-Sadler, S.J., Poole, D.M., Henrissat, B., Hazlewood, G.P., Clarke, J.H. & Gilbert, H.J. (1994). Evidence for a general role for high-affinity non-catalytic cellulose binding domains in microbial plant cell wall hydrolases. *Mol. Microbiol.* **11**, 375–382.
13. O'Neill, G., Goh, S.H., Warren, R.A.J., Kilburn, D.G. & Miller, R.C., Jr. (1986). Structure of the gene encoding the exoglucanase of *Cellulomonas fimi*. *Gene* **44**, 325–330.
14. Kraulis, P.J., *et al.*, & Gronenborn, A.M. (1989). Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. *Biochemistry* **28**, 7241–7257.
15. Xu, G.-Y., *et al.*, & Harvey, T.S. (1995). Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry* **34**, 6993–7009.
16. Johnson, P.E., Joshi, M.D., Tomme, P., Kilburn, D.G. & McIntosh, L.P. (1996). Structure of the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **35**, 14381–14394.
17. Klein, C. & Schulz, G.E. (1991). Structure of cyclodextrin glycosyltransferase refined at 2.0 Å resolution. *J. Mol. Biol.* **217**, 737–750.
18. Kubota, M., Matsuura, Y., Sakai, S. & Katsube, Y. (1991). Molecular structure of *B. stearothermophilus* cyclodextrin glucanotransferase and analysis of substrate binding site. *Denpun Kagaku* **38**, 141–146.
19. Sorimachi, K., Jacks, A.J., Le Gal-Coëffet, M.-F., Williamson, G., Archer, D.B. & 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.
20. Coutinho, P.M. & Reilly, P.J. (1994). Structural similarities in glucoamylases by hydrophobic cluster analysis. *Protein Eng.* **7**, 749–760.
21. Jacks, A.J., Sorimachi, K., Le Gal-Coëffet, M.-F., Williamson, G., Archer, D.B. & Williamson, M.P. (1995). ^1H and ^{15}N assignments and secondary structure of the starch-binding domain of glucoamylase from *Aspergillus niger*. *Eur. J. Biochem.* **233**, 568–578.
22. Le Gal-Coëffet, M.-F., Jacks, A.J., Sorimachi, K., Williamson, M.P., Williamson, G. & Archer, D.B. (1995). Expression in *Aspergillus niger* of the starch-binding domain of glucoamylase. Comparison with the proteolytically produced starch-binding domain. *Eur. J. Biochem.* **233**, 561–567.
23. Knegtel, R.M.A., *et al.*, & Dijkstra, B.W. (1995). Crystallographic studies of the interaction of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 with natural substrates and products. *J. Biol. Chem.* **270**, 29256–29264.
24. Larson, S.B., Greenwood, A., Cascio, D., Day, J. & McPherson, A. (1994). Refined molecular structure of pig pancreatic α-amylase at 2.1 Å resolution. *J. Mol. Biol.* **235**, 1560–1584.
25. Mikami, B., *et al.*, & Sacchettini, J.C. (1993). The 2.0 Å resolution structure of soybean β-amylase complexed with α-cyclodextrin. *Biochemistry* **32**, 6836–6845.
26. Svensson, B., Clarke, A.J. & Svendsen, I. (1986). Influence of acarbose and maltose on the reactivity of individual tryptophanyl residues in glucoamylase from *Aspergillus niger*. *Carlsberg Res. Commun.* **51**, 61–73.
27. Williamson, M.P., Le Gal-Coëffet, M.-F., Sorimachi, K., Furniss, C.S.M., Archer, D.B. & Williamson, G. (1997). The function of conserved tryptophans in the *Aspergillus niger* glucoamylase 1 starch binding domain. *Biochemistry*, in press.
28. Cygler, M., Wu, S., Zdanov, A., Bundle, D.R. & Rose, D.R. (1993). Recognition of a carbohydrate antigenic determinant of *Salmonella* by an antibody. *Biochem. Soc. Trans.* **21**, 437–441.
29. Sharff, A.J., Rodseth, L.E. & Quiocho, F.A. (1993). Refined 1.8 Å structure reveals the mode of binding of β-cyclodextrin to the maltodextrin binding protein. *Biochemistry* **32**, 10553–10559.

30. Knegtel, R.M.A., *et al.*, & Dijkstra, B.W. (1996). Crystal structure at 2.3 Å resolution and revised nucleotide sequence of the thermostable cyclodextrin glycosyltransferase from *Thermoanaerobacterium thermosulfurigenes* EM1. *J. Mol. Biol.* **256**, 611–622.
31. Tormo, J., *et al.*, & Steitz, T.A. (1996). Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J.* **15**, 5739–5751.
32. Dijkhuizen, L., Penninga, D., Rozeboom, H.J., Strokopytov, B. & Dijkstra, B.W. (1995). Protein engineering of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. In *Perspectives on Protein Engineering and Complementary Technologies.* (Geisow, M.J. & Epton, R., eds), pp. 96–99, Mayflower Worldwide Limited, Birmingham, UK.
33. Din, N., Gilkes, N.R., Tekant, B., Miller, R.C. Jr., Warren, A.J. & Kilburn, D.G. (1991). Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. *Bio/Technology* **9**, 1096–1099.
34. MacKenzie, D.A., Spencer, J.A., Le Gal-Coëffet, M.-F. & Archer, D.B. (1996). Efficient production from *Aspergillus niger* of a heterologous protein and an individual protein domain, heavy isotope-labelled, for structure-function analysis. *J. Biotechnol.* **46**, 85–93.
35. Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. & Sykes, B.D. (1995). ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* **5**, 67–81.
36. Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. & Wüthrich, K. (1983). Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**, 479–485.
37. Jeener, J., Meier, B.H., Bachmann, P. & Ernst, R.R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546–4553.
38. Kumar, A., Ernst, R.R. & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1–6.
39. Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R.R. (1988). Clean TOCSY for ^{1}H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110**, 7870–7872.
40. Braunschweiler, L., Bodenhausen, G. & Ernst, R.R. (1983). Analysis of networks of coupled spins by multiple quantum NMR. *Mol. Phys.* **48**, 535–560.
41. Mueller, L. (1987). P.E.COSY, a simple alternative to E.COSY. *J. Magn. Reson.* **72**, 191–196.
42. Bodenhausen, G. & Ruben, D.J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185–189.
43. Kay, L.E., Marion, D. & Bax, A. (1989) Practical aspects of 3D heteronuclear NMR of proteins. *J. Magn. Reson.* **84**, 72–84.
44. Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. & Bax, A. (1989) Three-dimensional heteronuclear NMR of ^{15}N-labeled proteins. *J. Am. Chem. Soc.* **111**, 1515–1517.
45. Otting, G., Widmer, H., Wagner, G. & Wüthrich, K. (1986). Origin of t_1 and t_2 ridges in 2D NMR spectra and procedures for suppression. *J. Magn. Reson.* **66**, 187–193.
46. Kördel, J., Skelton, N.J., Akke, M., Palmer, A.G., III & Chazin, W.J. (1992). Backbone dynamics of calcium-loaded calbindin D_{9k} studied by two-dimensional proton-detected ^{15}N NMR spectroscopy. *Biochemistry* **31**, 4856–4866.
47. Palmer, A.G., Rance, M. & Wright, P.E. (1991). Intramolecular motions of a zinc finger DNA-binding domain from *xfin* characterised by proton-detected natural abundance ^{13}C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.* **113**, 4371–4380.
48. Mandel, A.M., Akke, M. & Palmer, A.G., III (1995). Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246**, 144–163.
49. Stone, M.J., Chandrasekhar, K., Holmgren, A., Wright, P.E. & Dyson, H.J. (1993). Comparison of backbone and tryptophan sidechain dynamics of reduced and oxidised *Escherichia coli* thioredoxin using ^{15}N NMR relaxation measurements. *Biochemistry* **32**, 426–435.
50. Kjær, M., Andersen, K.V. & Poulsen, F.M. (1994). Automated and semiautomated analysis of homo- and heteronuclear multidimensional nuclear magnetic resonance spectra of proteins: the program PRONTO. *Methods Enzymol.* **239**, 288–307.
51. Brünger, A.T. (1992). *X-PLOR Version 3.1. A System for Crystallography and NMR*. Yale University, New Haven, CT, USA.
52. Bagby, S., Harvey, T.S., Eagle, S.G., Inouye, S. & Ikura, M. (1994). NMR-derived three-dimensional solution structure of protein S complexed with calcium. *Structure* **2**, 107–122.
53. Fletcher, C.M., Jones, D.N.M., Diamond, R. & Neuhaus, D. (1996). Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures. *J. Biomol. NMR* **8**, 292–310.
54. Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints — automated assignment of ambiguous crosspeaks and disulfide connectivities. *J. Mol. Biol.* **245**, 645–660.
55. Hyberts, S.G., Goldberg, M.S., Havel, T.F. & Wagner, G. (1992). The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736–751.
56. Imberty, A. & Perez, S. (1988). A revisit to the three-dimensional structure of B-type starch. *Biopolymers* **27**, 1205–1221.
57. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.