

Solution Structure of the Granular Starch Binding Domain of Glucoamylase from *Aspergillus niger* by Nuclear Magnetic Resonance Spectroscopy

Kay Sorimachi¹, Amanda J. Jacks¹, Marie-Françoise Le Gal-Coëffet^{1,2}
Gary Williamson², David B. Archer² and Michael P. Williamson^{1*}

¹Krebs Institute for
Biomolecular Research
Department of Molecular
Biology and Biotechnology
University of Sheffield
P.O. Box 594, Firth Court
Western Bank, Sheffield
S10 2UH, UK

²Department of Genetics and
Microbiology, Institute of
Food Research, Norwich
Research Park, Colney
Norwich NR4 7UA, UK

The solution structure of the granular starch binding domain (SBD) of glucoamylase 1 from *Aspergillus niger* has been determined by heteronuclear multidimensional nuclear magnetic resonance spectroscopy and simulated annealing. A total of 1092 nuclear Overhauser enhancement-derived ¹H-¹H distance constraints, 137 dihedral constraints and 86 hydrogen bond constraints were incorporated into an X-PLOR simulated annealing and refinement protocol. The family of calculated structures shows a well defined β -sheet structure consisting of one parallel and six antiparallel pairs of β -strands which forms an open-sided β -barrel. The root-mean-square deviation (rmsd) of 53 individual structures to the calculated average structure for the backbone atoms of residues excluding the N terminus and two mobile loops is $0.57(\pm 0.10)$ Å while the rmsd for backbone atoms in β -strands is $0.45(\pm 0.08)$ Å. Structural features of the SBD in solution are compared to the X-ray crystal structure of a homologous domain of cyclodextrin glycosyltransferase (CGTase) in the free and bound forms. Titration studies with two ligands, maltoheptaose and β -cyclodextrin, show the existence of two binding sites. Examination of the tertiary structures shows these two sites to be at one end of the molecule on opposite faces. The majority of residues showing the largest ¹H and ¹⁵N chemical shift changes are located in loop regions. Many residues implicated in binding, based on these changes, are similar in location to previously identified binding site residues in the crystal structures of CGTase. Overall, the shift changes are small indicating that the SBD does not undergo large conformational changes upon ligand binding.

© 1996 Academic Press Limited

Keywords: glucoamylase 1; starch binding domain; *Aspergillus niger*; NMR; solution structure

*Corresponding author

Abbreviations used: PDB, Protein Data Bank; 1CDG, crystal structure of CGTase/maltose complex (Lawson *et al.* 1994; PDB accession code, 1cdg); 1CGT, crystal structure of free CGTase (Klein & Schulz, 1991; PDB, accession code, 1cgt); 1D, one-dimensional; 2D, two-dimensional; 3D, three-dimensional; β CD, β -cyclodextrin; CGTase, cyclodextrin glycosyltransferase (1,4- α -D-glucan 4- α -D-(1,4- α -D-glucano)-transferase (cyclising), EC 2.4.1.19); DQF-COSY, double quantum filtered correlation spectroscopy; G1, glucoamylase 1 (1,4- α -D-glucan glucohydrolase, EC 3.2.1.3); HSQC, heteronuclear single quantum coherence; ³J_{N α} , coupling constant between backbone H^N and H ^{α} ; ³J_{N β} , coupling constant between backbone nitrogen and H ^{β} ; ³J _{$\alpha\beta$} , coupling constant between H ^{α} and H ^{β} ; MH, maltoheptaose; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; ppm, parts per million; P.E.COSY, primitive exclusive COSY; rmsd, root-mean-square deviation; SBD, granular starch binding domain of G1; <SBD>, ensemble of calculated structures of SBD; SBD_{av}, calculated average structure; SBD_{av-min}, the energy-minimised average structure; TOCSY, total correlation spectroscopy; TPPI, time-proportional phase incrementation.

Introduction

Glucoamylase 1 (G1; 1,4- α -D-glucan glucohydrolase, EC 3.2.1.3) from *Aspergillus niger* is an exo-acting enzyme which catalyses the conversion of starch and other polysaccharides to β -D-glucose by hydrolysis of α -D-glucosidic bonds. Three domains have been identified: the N-terminal catalytic domain (residues 1 to 470, 55 kDa), the C-terminal granular starch binding domain (SBD; residues 509 to 616, 12 kDa), and a short but bulky linker (residues 471 to 508, 13 kDa) which is heavily O-glycosylated at the abundant serine and threonine residues (Svensson *et al.*, 1983). The linker joins the two main domains giving the enzyme an overall dumbbell shape which has been observed clearly by scanning tunnelling microscopy (Kramer *et al.*, 1993).

There is an abundance of starch-degrading and related enzymes from a number of sources including animals, plants, bacteria and fungi (Svensson, 1988; Jespersen *et al.*, 1991). Of these, some exist as a single domain with catalytic function while others have multiple domains where one is a carbohydrate binding domain. Examples of the latter include the following: G1 (Svensson *et al.*, 1989), α and β -amylases of bacterial origin (Itkor *et al.*, 1990; Bahl *et al.*, 1991; Nanmori *et al.*, 1983; Kitamoto *et al.*, 1988) and cyclodextrin glycosyltransferase (CGTase; EC 2.4.1.19; Nitschke *et al.*, 1990; Lawson *et al.*, 1994) which have multiple domains including a catalytic and a starch binding domain; C_{ex}, a β -1,4-glycanase from *Cellulomonas fimi* which has an N-terminal catalytic domain and a cellulose binding domain separated by a proline/threonine-rich linker (O'Neill *et al.*, 1986) and chitinase A1 from *Bacillus circulans* WL-12 (Watanabe *et al.*, 1990) which has a catalytic domain and a chitin binding domain, separated by a fibronectin type III module. In all of these examples, the binding domains are functionally active in the absence of the catalytic domain. They do not affect catalysis of soluble substrates, but they display specific adsorption profiles towards insoluble substrates and dramatically enhance their degradation (Takahashi *et al.*, 1985; Gilkes *et al.*, 1988; Ståhlberg *et al.*, 1991; Watanabe *et al.*, 1994). Despite the apparently common role of these carbohydrate binding domains their precise mode of action is not known. Structural information is required in order to understand the manner in which the substrate interacts with the binding domain prior to catalysis. It is also of interest to determine the functional relationship between carbohydrate binding and catalysis.

Structural information is presently available for the catalytic domain of some of these hydrolytic enzymes. These include the X-ray crystal structures of CGTase complexed with acarbose (Strokopytov *et al.*, 1995), and free (Aleshin *et al.*, 1992, 1994a) and complexed (Harris *et al.*, 1993; Aleshin *et al.*, 1994b; Stoffer *et al.*, 1995) forms of

glucoamylase from *Aspergillus awamori* var. X100. However, very little structural work has been carried out on the non-catalytic domains of these enzymes. Solution structures of two cellulose binding domains have been determined by nuclear magnetic resonance (NMR) spectroscopy; type I cellulose binding domain of cellobiohydrolase I from *Trichoderma reesei* (Kraulis *et al.*, 1989) and type II cellulose binding domain of C_{ex} from *C. fimi* (Xu *et al.*, 1995).

In addition, crystal structures have been determined for free CGTase and the CGTase/maltose complex from *Bacillus circulans* and *Bacillus stearothermophilus* (Klein & Schulz, 1991; Kubota *et al.*, 1991; Lawson *et al.*, 1994). CGTases primarily convert starch to a mixture of α , β and γ -cyclodextrins by an intramolecular transglycosylation reaction. The enzyme consists of two domains of which the C-terminal one (E) is the putative granular starch binding domain. This domain is homologous to the SBD of G1 from *A. niger* showing approximately 37% amino acid sequence identity.

In this paper we present the solution structure of the SBD from *A. niger* G1. With details obtained from this study we aim to further our understanding of the structure/function relationship of the SBD of G1 and its role in the intact enzyme. The structural features examined include tryptophan residues which have already been implicated in the interaction with insoluble substrates in this domain (Clarke & Svensson, 1984). Tryptophan residues have also been implicated in other carbohydrate binding domains (Din *et al.*, 1994; Goto *et al.*, 1994). The possible location and role of the two binding sites on opposite faces of the protein are also discussed. We have made comparisons of our SBD solution structure to the crystal structures of free and bound CGTase since a high resolution structure of the SBD has not been published previously. Comparison of their secondary structures has already shown similarities of length and locations of β -strand secondary structure (Jacks *et al.*, 1995). This will form the basis for the investigation of the dynamics in solution for free and bound SBD. In addition, this will assist in the NMR resonance assignment of SBD mutants in order to probe the characteristics and implications of key binding site residues.

Results and Discussion

The amino acid residue numbering of the SBD used in this paper is the same as that used for intact G1 (Svensson *et al.*, 1983); thus, the SBD represents residues 509 to 616. Similarly, the residues defining the putative starch binding domain and their numbering for the free and bound forms of CGTase are the same as those reported for the intact protein in their respective crystal structure determinations (Klein & Schulz, 1991; Kubota *et al.*, 1991; Lawson *et al.*, 1994).

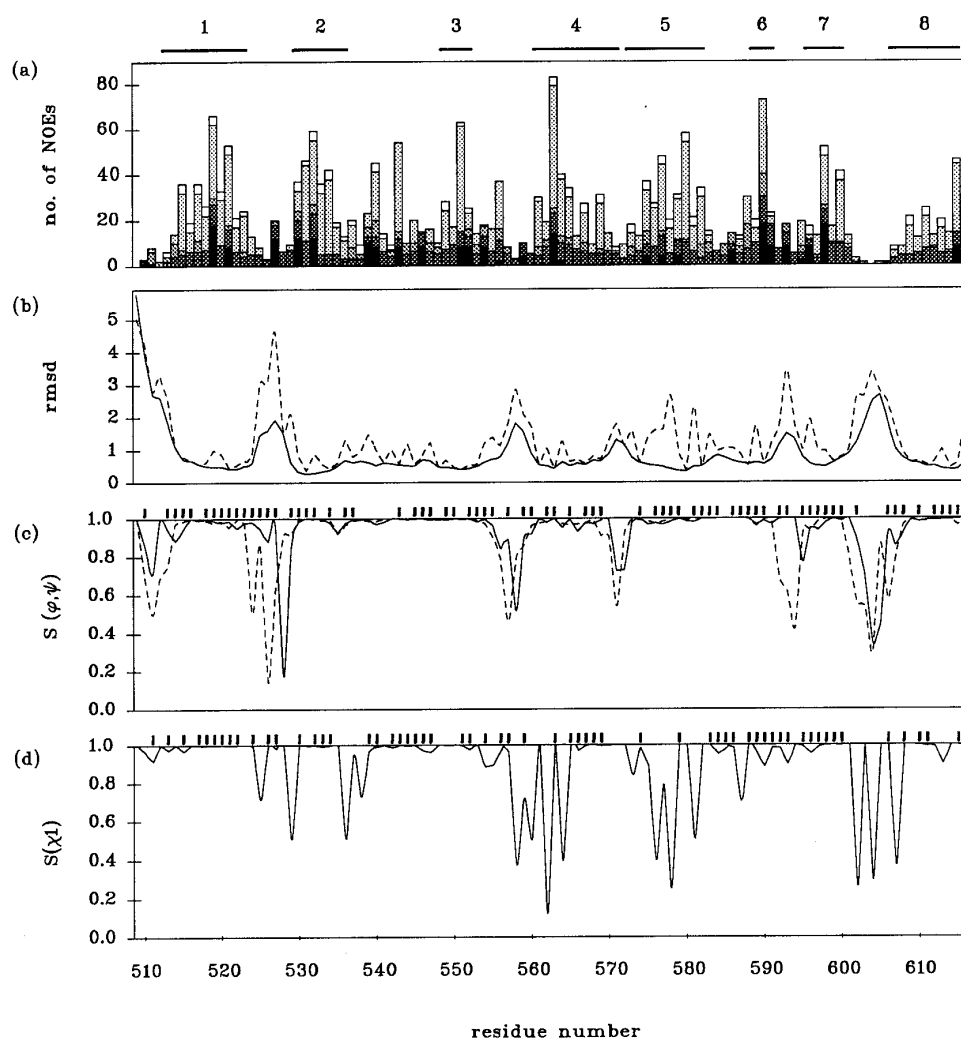


Figure 1. Summary of structural data *versus* residue number (indicated at the bottom of the Figure). The β -strand location and numbering are marked at the top. (a) Distribution of distance constraints. For each residue, bars representing the number of intraresidue (filled bar), sequential, medium-range, long-range NOEs and hydrogen bond (open bar) constraints are cumulatively stacked with decreasing intensity of shading. (b) Average rmsd (Å) of backbone (N, C $^{\alpha}$ and C; continuous line) and side-chain (broken line) atoms from SBD $_{av}$. (c) Angular order parameters for ϕ (continuous line), ψ (broken line), and (d) χ_1 dihedral angles for the 53 final structures. Residues whose ϕ and/or χ_1 angles were constrained in the structure calculation are marked above graphs (c) and (d), respectively.

Distance and dihedral angle constraints

^1H and ^{15}N NMR resonance assignments were made from two-dimensional (2D) homonuclear and 2D/three-dimensional (3D) heteronuclear data as described previously (Jacks *et al.*, 1995).

Distance and dihedral angle constraints were derived as described in Materials and Methods. A total of 115 intraresidue, 290 sequential, 100 medium-range, 587 long-range and 86 hydrogen bond distance constraints were obtained. Stereo-specific assignments were obtained for 35 out of 63 β -methylene pairs, seven out of eight valine γ -methyl groups and four out of six leucine δ -methyl groups. Dihedral angle constraints were obtained for 73 ϕ , 60 χ_1 and 4 χ_2 angles.

Structure determination

The family of structures for SBD was calculated using a simulated annealing protocol. Calculations were performed in an iterative fashion by checking violated constraints and adding new nuclear Overhauser enhancement (NOE) constraints after each stage. Ambiguous NOE assignments could often be resolved by inspecting the preliminary low resolution structures. In cases where NOE cross-peaks were found to be overlapped, the appropriate adjustment to the constraint was made. When the peaks could be partially resolved, a conservatively lower contour level was used. In more severely overlapped cases, the constraint was set to the weakest category. The final set of constraints incorporated in the calculation included 1092 NOE

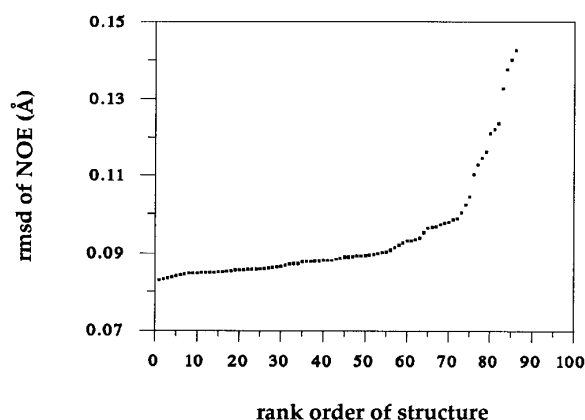


Figure 2. Values of rmsd of NOEs (Å) plotted against ranked order of structures. The values for structures 87 to 100 are greater than 0.15 Å.

distance constraints, 137 dihedral angle constraints and 86 hydrogen bond constraints, i.e. approximately 12 constraints per residue. The distribution

of NOE distance constraints is shown in Figure 1(a). A comparative lack of constraints is evident in three regions i.e. the N terminus and two loops spanning residues 523 to 529 and 601 to 606. These correspond to regions where resonance assignment was hampered by broadening/absence of signals (Jacks *et al.*, 1995).

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. Figure 2 shows the plot for rmsd of NOEs only, however, the shape of the curve was almost identical for all three parameters. The top 25% of structures with the highest rmsd values could be rejected immediately due to unreasonably high energies and a significant number of constraints that could not be satisfied. Within the remaining 75 structures, 53 had very similar energies and rmsd values and were selected to represent the family of solution structures of the SBD. The average

Table 1. Structural statistics and rmsd values for the family of 53 calculated structures of the SBD of *A. niger* G1

Parameters	$\langle \text{SBD} \rangle$	$\text{SBD}_{\text{av-min}}$
NOE violations >0.5 Å	$0.7(\pm 0.8)$	0
Dihedral violations $>5^\circ$	$3.2(\pm 1.1)$	1
rmsd from experimental restraints		
Distance restraints (Å)	$0.087(\pm 0.002)$	0.077
Dihedral restraints ($^\circ$)	$2.4(\pm 0.2)$	1.9
rmsd from idealised geometry		
Bonds (Å)	$0.0068(\pm 0.0002)$	0.0059
Angles ($^\circ$)	$1.00(\pm 0.03)$	0.90
Impropers ($^\circ$)	$0.91(\pm 0.04)$	0.84
X-PLOR energies (kcal mol^{-1}) ^a		
E_{TOT}	$1232(\pm 59)$	1009
E_{NOE}	$442(\pm 19)$	350
E_{CDIH}	$49(\pm 8)$	31
Atomic rmsd (Å)		
$\langle \text{SBD} \rangle$ versus SBD_{av}		
All	$1.12(\pm 0.29)$	$1.43(\pm 0.23)$
Residues (513-523, 530-600, 607-616) ^c	$0.57(\pm 0.10)$	$0.97(\pm 0.10)$
β -Strands 1-8 ^d	$0.45(\pm 0.08)$	$0.86(\pm 0.09)$
Buried residues ^e	$0.32(\pm 0.08)$	$0.79(\pm 0.12)$
SBD_{av} versus 1CGT ^{f,g}		
β -Strands 1, 2, 4-8	1.38	
β -Strands 1, 2, 4-7	1.13	
SBD_{av} versus 1CDG ^{f,g}		
β -Strands 1, 2, 4-8	1.42	
β -Strands 1, 2, 4-7	1.18	
1CGT versus 1CDG ^g		
β -Strands 1, 2, 4-8	0.30	

Structure notation: $\langle \text{SBD} \rangle$, ensemble of 53 calculated structures; SBD_{av} , calculated average structure; $\text{SBD}_{\text{av-min}}$, energy-minimised average structure.

^a Force constants used to calculate E_{NOE} and E_{CDIH} were $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, respectively, E_{TOT} represents total energy, E_{NOE} is the NOE energy term and E_{CDIH} is the dihedral angle energy term.

^b N, C $^\alpha$ and C atoms.

^c Excludes residues in N terminus (509 to 512) and loops between strands 1 and 2 (524 to 529), and between 7 and 8 (601 to 606).

^d β -Strand numbering and residues for SBD are as defined in this study.

^e Residues with surface accessibility $<3\%$ (see the text).

^f 1CGT and 1CDG refer to the crystal structures of free CGTase (Klein & Schulz, 1991) and CGTase/maltose complex (Lawson *et al.*, 1994), respectively.

^g To compare structures, the maximum number of aligned β -strand residues common to all three (SBD, 1CGT and 1CDG) structures were used in the superimposition.

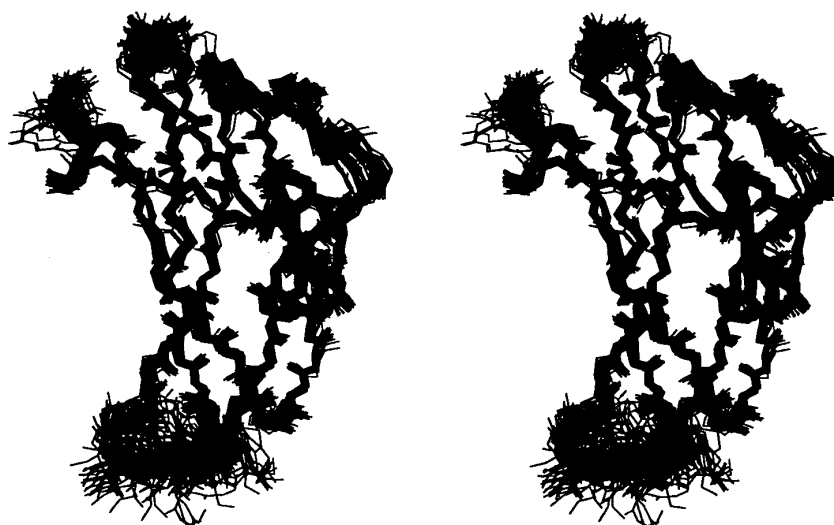


Figure 3. Stereo view of the backbone (N, C α , C and O) atom superimposition of 53 calculated structures of the SBD of G1 from *A. niger*. The structures were superimposed to SBD_{av} on the N, C α and C atoms of the eight β -strands (residues 513 to 523, 530 to 536, 549 to 552, 561 to 571, 573 to 582, 589 to 591, 596 to 600, 607 to 615). The N and C termini are at the bottom and top of the Figure, respectively.

structure (SBD_{av}) was calculated from the 53 selected structures and was further subjected to restrained energy minimisation to yield the minimised average structure (SBD_{av-min}).

Structural statistics for the calculated ensemble are given in Table 1. Figure 3 shows a backbone atom superimposition of the 53 structures to SBD_{av}. The atomic rmsd for backbone and all heavy atoms when superimposed over the eight β -strand residues are $0.45(\pm 0.08)$ Å and $0.86(\pm 0.09)$ Å, respectively. When all residues are used in the superimposition, the atomic rmsd for backbone and all heavy atoms are $1.12(\pm 0.29)$ Å and $1.43(\pm 0.23)$ Å, respectively. When residues in the N terminus and the loops between β -strands 1 and 2 and between β -strands 7 and 8 (a total of 17 residues out of 108) are eliminated from the superimposition, the backbone and all heavy atom rmsd values are reduced to $0.57(\pm 0.10)$ Å and $0.97(\pm 0.10)$ Å, respectively. These values are close to those obtained when only the β -strand residues (57 in total) are used in the superimposition and indicates that the other loops, including the longest one of 12 residues between strands 2 and 3, are reasonably well defined. This is reflected by the large numbers of constraints in these regions (Figure 1(a)). Superimposition of backbone and all heavy atoms of residues with low solvent accessibility (discussed below) gives rmsd values of $0.32(\pm 0.08)$ Å and $0.79(\pm 0.12)$ Å, respectively, as would be expected of buried residues.

Description of solution structure

The overall topology of the SBD of G1 from *A. niger* shows eight β -strands forming two major β -sheets (Jacks *et al.*, 1995). One sheet consists of five strands arranged in antiparallel fashion while

the second sheet has one parallel and one antiparallel strand pair (Figure 4). The N and C termini are at opposite ends of the longest axis of the molecule. Analysis of structures and hydrogen bonding patterns shows that the residues involved in β -strands in order of sequence are as follows: 513 to 523, 530 to 536, 549 to 552, 561 to 571, 573 to 582, 589 to 591, 596 to 600 and 607 to 615. Figure 5 gives a representation of the β -strands. Residues A523,

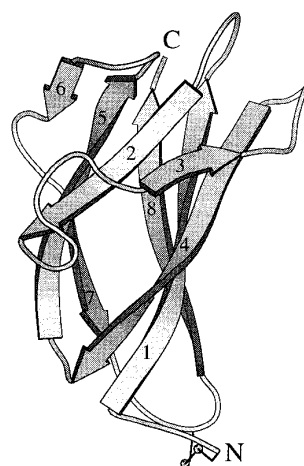


Figure 4. Representation of the solution structure of the SBD. The N and C termini and β -strand numbers are marked. The strands are shown as arrows which also indicate their directionality. The atomic coordinates of SBD_{av-min} were used and the molecule was oriented by eye to give the best view of the β -strands. The orientation is rotated by approximately 180° about the vertical axis compared to Figure 3. The Figure was generated using the program MOLSCRIPT (Kraulis, 1991). The position of the disulphide bond is highlighted by ball figures for S γ atoms and lines for C α -C β , C β -S γ (both intraresidue) and S γ -S γ (interresidue) bonds for residues 509 and 604.

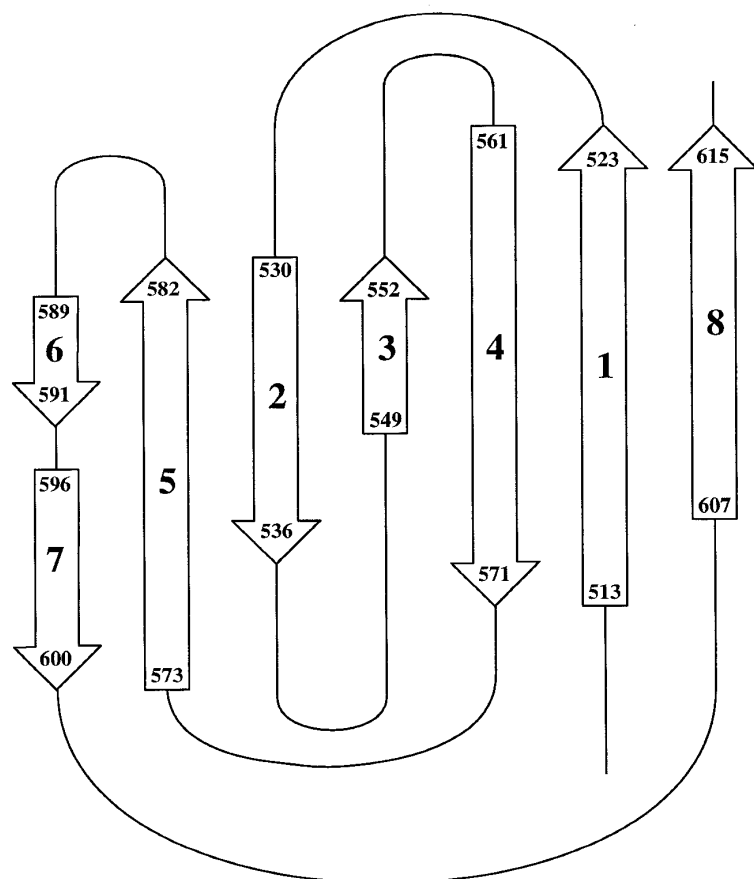


Figure 5. Representation of the direction and alignment of the β -strands (numbered 1 to 8) in SBD. The first and last residues in each strand are marked at the ends of the arrows.

P561 and L562 represent additional β -sheet residues to those previously identified (Jacks *et al.*, 1995) due to an extension of the antiparallel interaction of strands 1 and 4. The β -strands are well stabilised as evidenced by the slow amide exchange rates for these residues, which in some cases have half-lives of weeks or months at 313 K and pH 5.2, and form the core of the SBD. The longer strands (1, 4, 5 and 8) show a degree of curvature and twist which allows better packing. In addition we have now identified two hydrogen bonds (discussed below) between strands 3 and 4 involving residues S552 and Y564 in antiparallel fashion. Thus the SBD forms an open-sided, distorted, β -barrel structure. There are six loops of significant length, four of which are well defined. The N terminus is close to the loop between strands 7 and 8 and is connected to it by a disulphide bond which links C509 and C604. The approximate dimensions of the SBD are $42 \text{ \AA} \times 38 \text{ \AA} \times 31 \text{ \AA}$.

Analysis of structures

Figure 6 shows the distribution of NOE constraints used in the structure calculation and the NOE violations resulting from it. Interactions involving at least one backbone proton and those involving only side-chain protons have been separated into the two halves of the diagonal plots. In (a), groups of constraints are evident for each

β -strand. Most of these run perpendicular to the diagonal since the strands are arranged mainly in antiparallel fashion. The constraints are more clustered when they involve a backbone proton (lower left of Figure 6) since most of these are due to cross-strand NOEs and occur in an ordered pattern. In (b), the NOE violations are shown as the average violation ($\text{\AA}$) per constraint. In structure determinations, the quality of calculated structures is usually expressed in the form of various statistical data. For the SBD, these are presented in Figure 1 and Table 1 which indicate that the calculated structures show good convergence and precision. An aspect often neglected from such an analysis is the NOE violations, apart from the usual practice of reporting the average number of violations per structure that fail to satisfy an arbitrarily chosen cutoff distance. We have presented Figure 6(b), which analyses NOE violations resulting from our calculations, with two main aims. The first is to provide a readily apparent overview of the extent and location of NOE violations. It is clear from Figure 6 that many constraints are satisfied (i.e. boxes present in (a) are absent in (b)), and that there is no region that is markedly worse than any other. Thus, we have a workable constraint set with a good and essentially completely self-consistent distribution of constraints. The second aim is to highlight the most significant violations that may be worth studying in

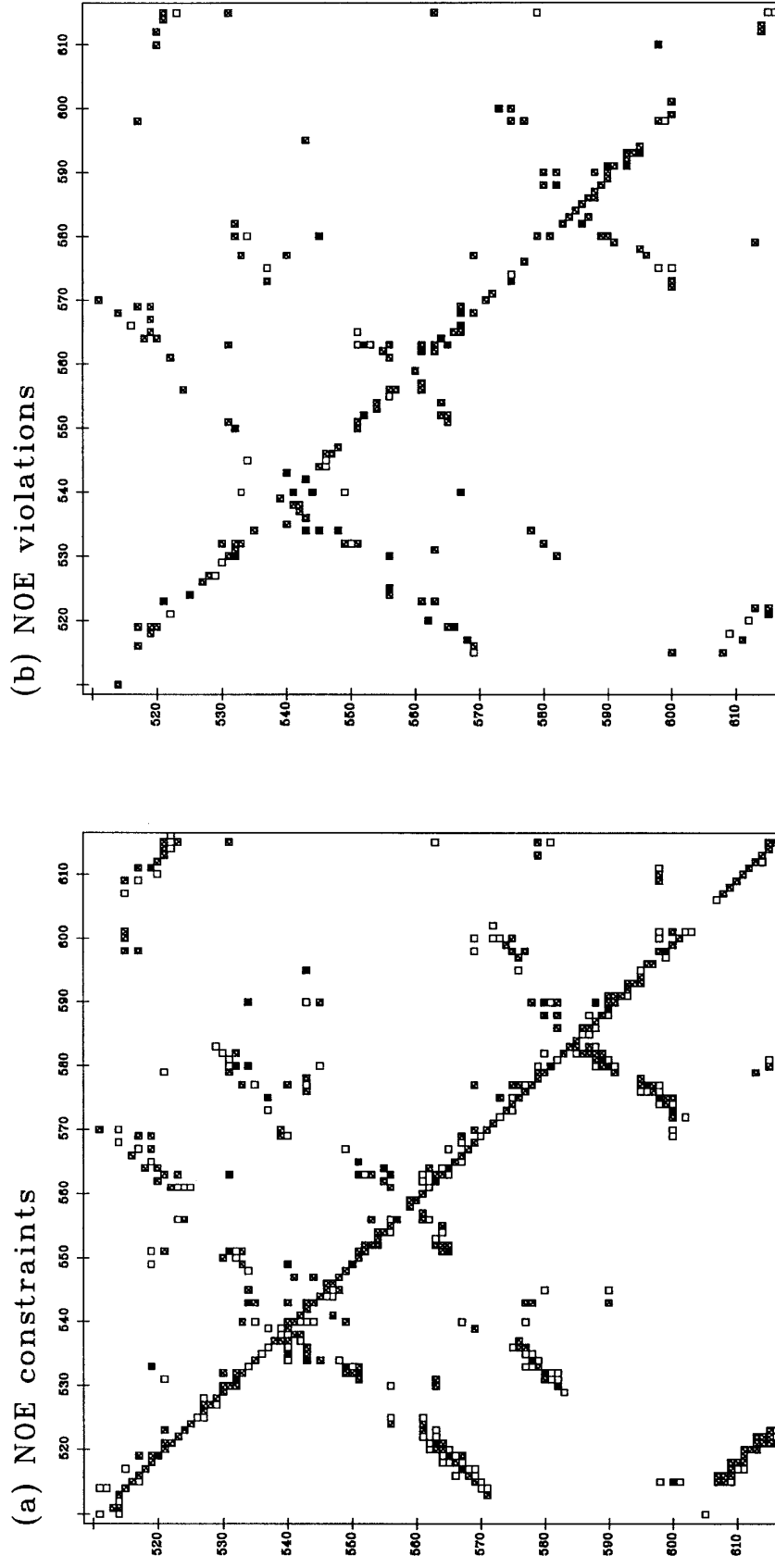


Figure 6. Diagonal plots for SBD_{av-min}, showing backbone to backbone and backbone to side-chain interactions (lower left) and side-chain to side-chain interactions (upper right). Residue numbering is shown on the left and top of the axes. (a) Total number of NOE constraints used: 1, open squares; 2 to 4, hatched squares; ≥ 5 , filled squares. (b) Severity of NOE violations shown as the average violation (Å) per constraint: ≤ 0.01 , open squares; > 0.01 but ≤ 0.1 , hatched squares; > 0.1 , filled squares. Where a constraint is not violated, no square is drawn.

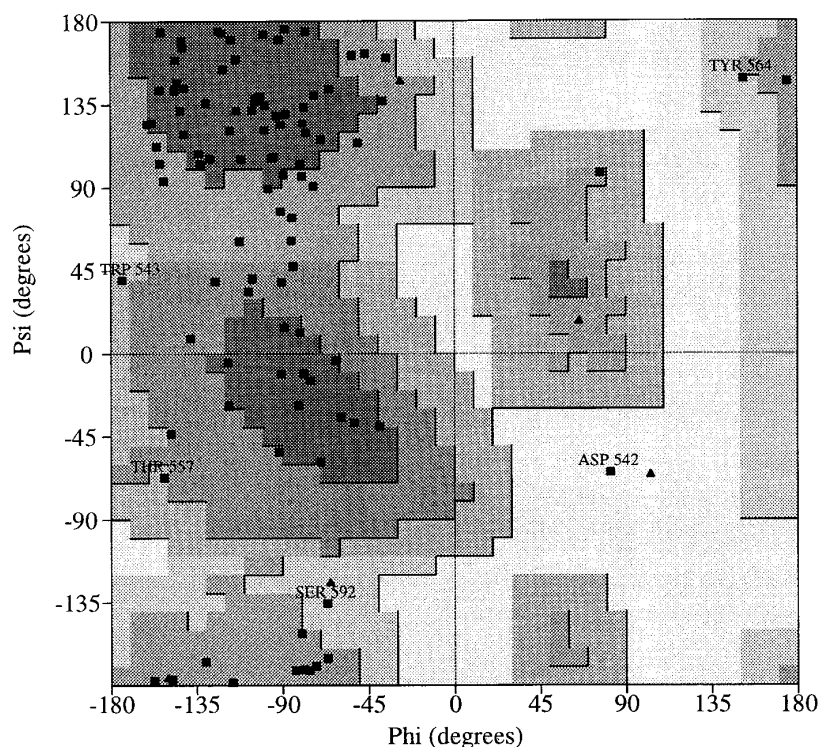


Figure 7. Ramachandran plot of ϕ and ψ dihedral angles for SBD_{av-min}. The plot was generated using the program PROCHECK v.3.0 (Laskowski *et al.*, 1993). The shaded areas indicate, in order of decreasing intensity, most favoured region, additional allowed region, generously allowed region, and disallowed region. Residues in the latter two regions are labelled. Glycine residues are shown as triangles.

more detail. A violation of an NOE constraint indicates that the constraint could not be satisfied; this may be due to a convergence problem in the simulated annealing protocol, or it may indicate a more serious problem such as an incorrectly assigned NOE or a constraint which is too tight. Errors of the latter type will act to distort the structure and must therefore be excluded at all costs. Constraints that are part of a dense cluster of other local constraints have little distorting effect because the other constraints will act to maintain the local geometry correctly. The most serious distortions will occur where there are isolated NOEs that are in error. Such NOEs will show up in Figure 6(b) as isolated shaded boxes, such as that between L540 (H^N) and V567 (H^γ) on the lower left. This method allowed us to quickly identify potentially crucial errors which can easily become lost in very long constraint lists. Thus, the L540 H^N -V567 H^γ NOE was verified and the 0.191 Å violation of this constraint in SBD_{av-min} can be attributed to a convergence problem.

In Figure 1(a) the distribution of distance constraints per residue is shown while in Figure 1(b) the backbone and side-chain rmsd values of SBD_{av} have been plotted. A general trend can be observed where the loop regions have fewer constraints, which can be attributed to greater mobility and incomplete assignment. Consequently the structure is not as well defined in these regions and gives rise to higher rmsd values. Figure 1(c) and (d) are plots of the angular order parameters (S) of ϕ , ψ and χ_1 dihedral angles. An S value approaching unity reflects that the local geometry is well defined in the 53 calculated structures. This

is generally the case for strand residues where both backbone and side-chain are ordered and also for the backbone of residues in the loop region between strands 2 and 3. Low $S(\psi)$ values are often followed by low $S(\phi)$ values because variability in the orientation of the peptide bond between residues i and $i+1$ produces concerted changes in $\psi(i)$ and $\phi(i+1)$. These graphs show that dihedral angle constraints are extremely powerful in their ability to define local geometry since low S values generally occur where such constraints were not or could not be employed. We also calculated three-residue rmsd values (data not shown) which revealed a similar trend to the $S(\phi, \psi)$ plot (Figure 1 (c)).

Figure 7 is a Ramachandran plot of ϕ and ψ dihedral angles in the SBD_{av-min} structure. Residue D542 is the single non-glycine residue occurring in a disallowed region of the plot and possesses a ϕ angle of 81.9° and a $^3J_{N\alpha}$ value of 6.2 Hz. This residue is found within a network of hydrogen bonds (discussed below). To test that the D542 H^N -S538 O constraint was not the source of this unusual backbone geometry, a restrained minimisation without this distance constraint was performed. The result showed that the loop conformation in this region is well defined and the ϕ angle of D542 did not change significantly. Examination of χ_1 - χ_2 plots (not shown) indicate that most residues adopt favourable or allowable side-chain conformations.

The β -strands of the SBD are very well defined. Characteristic hydrogen bond and cross-strand NOE patterns are observed in most strands (shown in Figure 5 of Jacks *et al.*, (1995)). Most of the

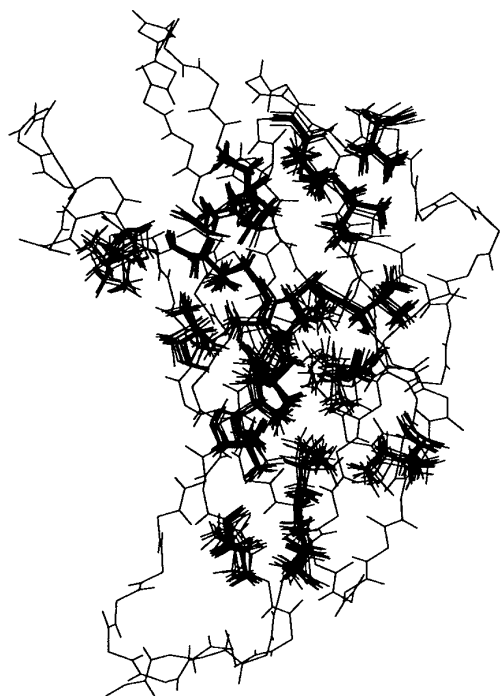


Figure 8. Distribution of non-aromatic branched hydrophobic residues in the SBD structure. The backbone (N, C α , C) atoms of ten final structures selected at random from the family of 53 structures were superimposed onto SBD_{av-min}. The backbone (N, H α , C α , C, O) atoms of SBD_{av-min} and side-chain atoms of valine, leucine and isoleucine residues for the ten calculated structures are displayed. The orientation of the molecule is the same as in Figure 3.

branched hydrophobic residues (valine, leucine and isoleucine residues) are found in the strands of the molecule. The exceptions out of 19 such residues are I537, L540, L562 and V588; of these, L540 is the only one which is not located immediately adjacent to a strand. The distribution of hydrophobic residues (Figure 8) shows that they form the core of SBD. A total of 21 residues with less than 3% surface accessibility were regarded as buried; these are V515, V517, F519, L521, A523, I531, L533, V534, G535, L540, L551, V567, L569, F575, Y577, K578, I580, I582, V600, V611 and D613. The majority of these residues are hydrophobic in nature as would be expected. The only polar residues in this list are K578 and D613; both of these have access to solvent through the ends of their side-chains.

The majority of residues found in the loop regions are charged or hydrophilic consisting of Asp, Glu, Lys, Ser or Thr residues. The long loop between strands 2 and 3 is surprisingly well defined owing to the large number of interresidue NOEs involving, in particular, residues L540 and W543, as well as eight χ_1 angle constraints. Following preliminary calculations, hydrogen bond pairs were identified in this region as discussed below. These add significantly to the definition of structure in this loop which is reflected by the low

backbone, side-chain and three-residue rmsd values and angle order parameters close to unity (Figure 1). The least well defined region of the structure is the bottom end of the protein as shown in Figure 4, consisting of the N-terminal residues 509 to 512 and the loop residues 601 to 606. These regions are linked by a disulphide bridge between C509 and C604 (Williamson *et al.*, 1992). In the full length protein the catalytic domain is attached at C509 via a glycosylated linker region. Thus, in the intact protein the dynamics of this region could be quite different. However, if it were the case that this region is still flexible in the intact protein, the flexibility could be functionally significant, in that it implies that the catalytic domain can access a large area of starch around the binding site, by "hinging" on the flexible region around C509.

On the basis of hydrogen bonding patterns and NOEs, we previously proposed that there is a β -bulge in strand 3 (Jacks *et al.*, 1995). The tertiary structure calculation shows that this is not present. Examination of the calculated structures shows that residue S552 adopts a conformation such that the backbone amide and carbonyl groups are pointing away from strand 2. In these circumstances it is not possible for two hydrogen bond pairs deduced in our previous study (Jacks *et al.*, 1995), S552 H α -I531 O and I531 H α -S552 O, to form. The H α of S552 is now found to be hydrogen bonded to Y564 O instead. This is corroborated by a Y564 H α -S552 O hydrogen bond pair. These two interactions provide a non-sequential link between the two major β -sheets thus forming a β -barrel which has an open end, since there is no evidence of any interaction between strand 8 and strands 6 and 7.

Some protons with slow or medium exchange rates were previously identified (Jacks *et al.*, 1995) but their possible hydrogen bonding partners could not be determined with confidence due to lack of supportive NOE data. Several of the partners have now been identified. Firstly, the pair A523 H α -P561 O can be seen as an extension of the interaction between strands 1 and 4. The amide proton of G535 is in slow exchange but previously there was no obvious hydrogen bond acceptor since it cannot interact with strands 3 and 5 due to its orientation and the length of these strands. We have now identified the oxygen atom of L540 as its proton acceptor and this hydrogen bond contributes to the stabilisation of the long loop between strands 2 and 3. Other hydrogen bond pairs identified and contributing to the network in this region are L540 H α -I537 O, G541 H α -S538 O and D542 H α -S538 O. Many NOEs are also found in this region which explains why this loop is very well defined. The hydrogen bond pair between E583 H α and S587 O occurs in the loop connecting strands 5 and 6. This restricts the type of motion where a loop may flip or bend at its base towards the flat face formed by the pair of strands. We also anticipated a pair between E576 H α and S536 O since we have observed the S536 H α -E576 O hydrogen bond. However, the backbone dihedral angle of S536

Table 2. Observed chemical shift changes ($\Delta\delta$) of SBD ^1H and ^{15}N resonances upon titration with βCD or MH

Secondary structure element ^a	Residue involved ^b	^1H (ppm)	$\Delta\delta^c$ ^{15}N (ppm)
S1	L521	-0.040 ^d (-0.059) ^d	
L1/2	T524	0.055 ^e	
	Y527	0.078	
	E529	0.048	
S2	N530	0.041	
	I531	-0.117 ^f	
L2/3	L540	-0.089 ^e	
	W543	0.119 ^d (0.131) ^d	
L3/4	E544		0.335
	D554	-0.052	-0.356
	T557	0.058	0.491
S4	L562	-0.073 ^f	
	W563	0.086	
	Y564	0.104 ^e	
S5	E576	-0.059 ^{d,e} (-0.034) ^d	
	I582	0.043	
L5/6	S584	0.084	
S6	E591	-0.044	
L6/7	S592	0.090	0.325
	D593	0.043	
	N595		-0.406

^a Secondary structural features determined in this study: Sx = β -strand number x; Ly/z = loop between strands y and z.

^b SBD residues which are homologous to binding site residues in the CGTase/maltose complex (Lawson *et al.*, 1994) are shown in boldface. Residues proposed as being involved in interactions with substrate from hydrophobic cluster analysis (Coutinho & Reilly, 1994b) are shown in italics.

^c Data shown for $\Delta\delta > 0.04$ ppm for ^1H and $\Delta\delta > 0.3$ ppm for ^{15}N . All data were obtained from βCD titrations except for the values in parentheses which were from MH titrations and are shown under the corresponding βCD titration value. $\Delta\delta$ was monitored by 2D HSQC spectra, unless otherwise stated.

^d Monitored by 1D ^1H spectra.

^e Monitored by 2D TOCSY experiments.

^f Results obtained from HSQC and TOCSY spectra.

($\phi = -89^\circ$, $\psi = 14^\circ$) does not allow the formation of the former hydrogen bond as the backbone chain twists away from strand 5. A similar situation is encountered with the backbone of strand 6 at residue E591 and hence a plausible acceptor for F579 H^{N} was not identified. Inspection of the structure shows that part of the aromatic ring of F579 is exposed but the amide group is inside a cavity. Suitable hydrogen bond acceptor atoms to the backbone amide protons of I531 and S574 and side-chain H^{e1} of W563 were also not identified. In the cases of I531 and W563, the calculated surface accessibility for these residues were 1.2% and 4.3%, respectively; therefore they are likely to be buried sufficiently to retard their exchange with solvent. Residue I531 in particular is totally surrounded by aromatic groups including W563, W615 and Y532 at close range and F579 slightly further away. This is reflected in the upfield chemical shift of I531 H^{N} at 5.48 parts per million (ppm) (Jacks *et al.*, 1995).

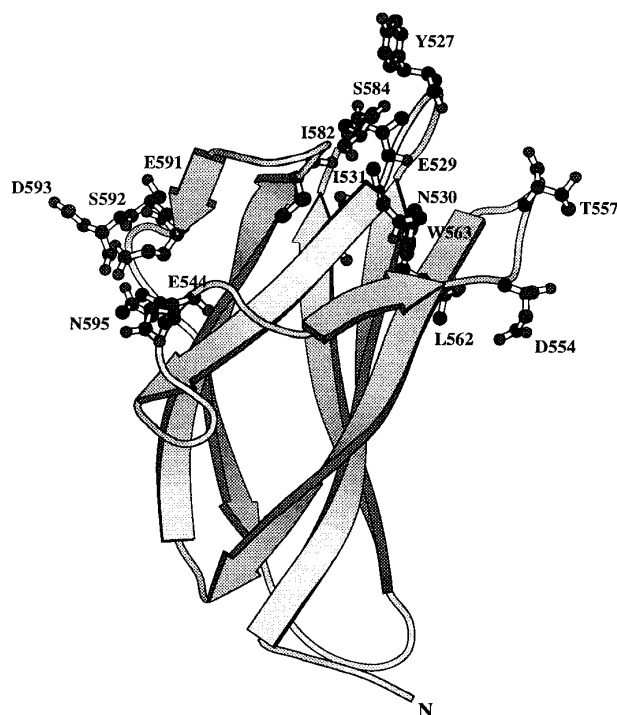


Figure 9. Representation of the SBD showing residues affected in HSQC titration experiments with βCD . Residues whose chemical shift of the backbone or side-chain nitrogen atom showed a change >0.3 ppm and/or the ^{15}N -bound proton chemical shift change was >0.04 ppm are labelled and represented as ball-and-stick. The atomic coordinates of SBD_{av-min} were used and the molecule orientation is the same as Figure 4. The Figure was generated using the program MOLSCRIPT (Kraulis, 1991).

Characteristics of ligand binding

A number of experiments were performed to analyse the effects of ligand binding to the SBD. Chemical shift changes were measured for selected resolved resonances of the SBD with the addition of maltoheptaose (MH) which is a linear substrate for the SBD and β -cyclodextrin (βCD), a cyclic analogue of starch. The results are shown in Table 2 as the overall change per residue for ^1H shift changes >0.04 ppm and ^{15}N shift changes >0.3 ppm. Average values were determined if chemical shift changes were obtained for more than one proton in a particular residue.

The majority of residues associated with these changes are located in loop regions closer to the C-terminal end of the molecule (the top half of the protein in the orientation shown in Figure 4). As seen for residues L521, W543 and E576, the shift changes were found to be of similar orders of magnitude for the two ligands, in agreement with Kusnadi *et al.* (1994). The largest net shift change for ^1H resonances was associated with residue W543 and for ^{15}N resonances, residue T557. Figure 9 highlights the residues whose chemical shifts are affected (according to the above criteria) in the

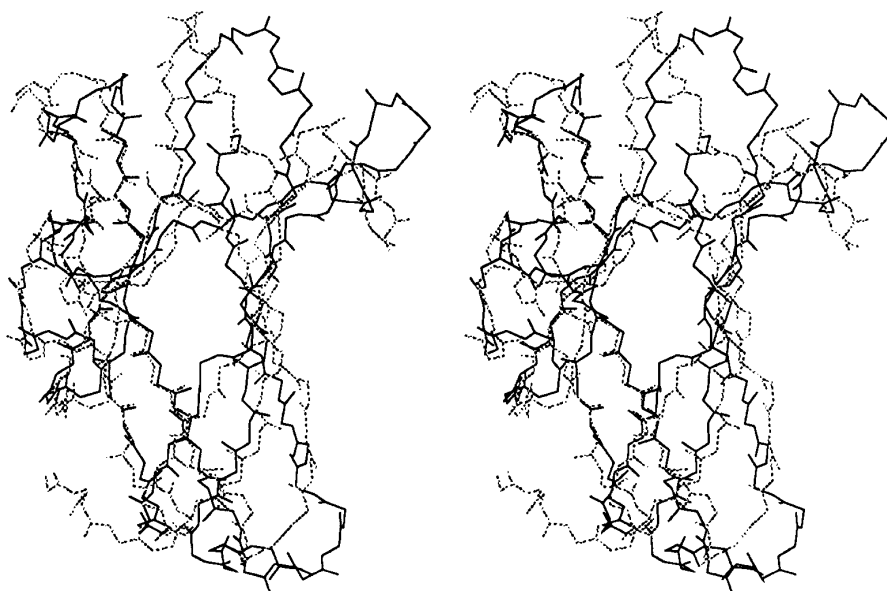


Figure 10. Stereo view of the backbone (N, C α , C and O) atom superimposition of SBD_{av-min} (continuous line) and the crystal structure of 1CGT (broken line). The structures were superimposed on the N, C α and C atoms of β -strands 1, 2, 4 to 7 of SBD_{av-min} and the equivalent residues in the CGTase structure.

heteronuclear single quantum coherence (HSQC) titration experiments.

Many residues affected by the titration are homologous to those in CGTase which have been identified as binding site residues (discussed below). Others surrounding these titrated residues are also affected due to their proximity to these residues or possibly due to a conformational change of the protein. However, since only moderate chemical shift changes have been observed overall, and the affected residues are limited to two defined regions of the protein, any conformational change associated with the protein is expected to be conservative. The comparatively large ^1H shift change associated with I531 is surprising since it has very low solvent accessibility and therefore is not expected to contact the ligand directly. However, residue W563, a potential binding site residue (see later), is also affected by the titration. Ring current shift calculations (Williamson & Asakura, 1993) on SBD_{av-min} indicate that the resonance frequency of I531 H^N is shifted 1.7 ppm upfield by W563. Thus, a small change in the conformation of W563 could have a large effect on the chemical shift of I531.

Comparison with crystal structures of free and bound CGTase

In the absence of other high resolution structures of the SBD of G1, we have made structural comparisons of our calculated solution structure to crystal structures of the homologous domain (E) in CGTase in the free form (1CGT; Klein & Schulz, 1991) and complexed to maltose (1CDG; Lawson *et al.*, 1994). Note that there are small

differences in the sequence between 1CGT and 1CDG arising from the different bacterial strains of the enzyme.

Overall topology

As previously reported (Jacks *et al.*, 1995), both SBD and CGTase structures are β -sheet in character and the location and length of β -strands are very similar. This is evident in Figure 10 where the structures of SBD_{av-min} and 1CGT were superimposed. This is despite the partial β -barrel structure of SBD which we observe resulting from the identification of new hydrogen bond interactions between strands 3 and 4. There is no evidence for such an interaction in CGTase which adopts a different conformation at the end of the third strand. Other differences are observed in the N terminus and four loop regions spanning residues 524 to 529, 553 to 560, 583 to 588 and 601 to 606 of SBD. In most cases, the variation is greatest in the central part of these loops. The vastly different conformation of the N terminus (at the bottom of Figure 10) is primarily attributable to the disulphide linkage (C509-C604) in SBD which is absent in CGTase. This region is protruding more in SBD while the N terminus of CGTase points away from the domain in both the free and bound (not shown) forms. This observation, however, is not likely to affect substrate binding or the activity of the intact enzyme since firstly, the binding sites are on the other side of the molecule (see later discussion) and secondly, in G1 the SBD is connected through the N terminus to a glycosylated linker and the catalytic domain.

Local structural differences

In the core of SBD, β -strands 1, 2 and 4 to 7 are very similar to the homologous strands in CGTase. Strand 8 shows the greatest displacement which is due to the difference in length (two residues longer in SBD). The *cis*-proline residue in strand 3 of CGTase produces a kink in the backbone which differs from its conformation in SBD. Residue T518 of SBD (strand 1) is an arginine in CGTase whose side-chain protrudes much further out, away from the strand and the molecule. The implication of this (if any) is not immediately obvious but in CGTase, R588 may provide an interdomain interaction to N203, thus contributing to stability in the intact enzyme. Two of the aromatic residues in strand 5, Y577 and F579, have their phenyl rings in a different orientation to those of F649 and F651, respectively, of 1CGT. In both cases, the rings are approximately on opposite sides of the C^α - C^β axis.

Many *cis*-proline residues were found in the structure of 1CGT (Klein & Schulz, 1991). In SBD, NOE data were consistent with all proline residues adopting a *trans* conformation (Jacks *et al.*, 1995) except that residues P512 and P570 showed evidence of *cis-trans* isomerisation. It is noteworthy that P512 and P570 are in adjacent regions of the structure, at the lower end of the protein (in Figure 4), a region identified as the most disordered part of the protein. These residues are located before the beginning of β -strand 1 and at the end of strand 4, respectively (see Figure 5). The sharp loop in 1CGT containing *cis*-P633 is stabilised by two hydrogen bonds, A594 H^N -P633 O and T634 H^N -Q631 O (Klein & Schulz, 1991). A hydrogen bond analogous to the first pair is observed in SBD as A523 H^N -P561 O which is in fact a cross-strand interaction since we have identified A523 and P561 to be part of β -strands while the equivalent residues in CGTase are in loop regions. In strand 3, the backbone conformation of SBD is different to that in 1CGT possibly due to *cis*-P623 in the latter; however, the side-chain of A550 of SBD occupies similar conformational space to the proline ring.

The four conserved tryptophan residues are of interest since some of these have been implicated in substrate binding (Clarke & Svensson, 1984; Svensson *et al.*, 1986). The side-chain of residues W543 and W563 of SBD are in almost identical location and orientation to W614 and W635, respectively, of 1CGT. The indole ring of W590 is tilted slightly relative to W661 of 1CGT but they occupy similar overall conformational space. Due to the slight difference in conformation of strand 8 the location of W615 of SBD shows the largest variation relative to W683 of 1CGT. However this is not unexpected since in Figure 10, strand 8 was

omitted from the superimposition and also being in the C terminus we may anticipate some degree of flexibility. Residues W543 and W590 are most exposed to solvent and their indole ring surfaces are almost adjacent in the SBD structure. The equivalent residues in CGTase structures (W614 and W661, respectively, in 1CGT; W616 and W662, respectively, in 1CDG) have both been identified as binding site residues. Of the four tryptophan residues in SBD the above two would be the most likely candidates for substrate binding in terms of accessibility.

Although residue D542 exhibits unusual backbone geometry as discussed earlier, the conformation adopted by this and adjacent residues is well defined. This is evident in Figure 1 which shows low rmsd values and high (>0.98) angular order parameters for all three dihedral angles reflecting well ordered local geometry. Analysis of the backbone dihedral angles of residues homologous to D542 of SBD in free CGTase (N613) and the CGTase/maltose complex (N615) reveals that both asparagine residues possess a positive ϕ angle although they are outside the disallowed area of a Ramachandran plot surface. The location of the side-chain for these residues differs slightly between SBD and CGTase, however, the C^β - C^γ - $O^{\delta 1}$ - $O^{\delta 2}$ surface and C^β - C^γ - $O^{\delta 1}$ - $N^{\delta 2}$ surface, respectively, lie in the same plane revealing a similar electronegative surface. This unusual geometry may be necessary to allow the indole ring of the adjacent residue W543 (a putative binding residue) to adopt the correct orientation for interaction with the substrate.

Possible binding site residues

The substrate binding sites in domain E of CGTase have been studied previously in detail by Lawson *et al.* (1994) in the crystal structure of the CGTase/maltose complex. In addition, starch binding residues in SBD have been proposed on the basis of sequence alignment of CGTase and glucoamylases from various sources and hydrophobic cluster analysis (Coutinho & Reilly, 1994a,b). In order to gain a better understanding of the possible binding sites in SBD, some of the observed interactions in 1CDG were compared to the SBD solution structure determined in this study.

The interaction of maltose with domain E of CGTase in 1CDG (Lawson *et al.*, 1994) is observed at two distinct sites. Site 1 $\dagger$ involves five residues (W616, K651, W662, E663 and N667) while site 2 involves seven residues (T598, A599, G601, N603, N627, Q628 and Y633). In the structure of *B. stearothermophilus* CGTase/maltose complex, Kubota *et al.* (1991) previously identified T591, N596, N620, Y626 and W629 as substrate binding site residues. These mostly correspond to site 2 residues in 1CDG and no mention is made of the existence of a second site. Binding site residues of SBD identified by hydrophobic cluster analysis

$\dagger$ The numbering of binding sites used by Klein & Schulz (1991) and Coutinho & Reilly (1994b) is opposite to that of Lawson *et al.* (1994). We have used the latter definition since most of our structural comparisons are made to 1CDG.

(Coutinho & Reilly, 1994b) using knowledge of the SBD disulphide bridge, CGTase binding site residues and homology between SBD and CGTase residues amongst other criteria, gave a more generous definition. The proposed SBD residues were for site 1, W543, E576, K578, W590, N595 and for site 2, T525, T526, Y527, G528, E529, N530, D554, K555, D560, W563.

Modelling the binding sites in 1CDG from inspection of electron density maps shows that the maltose molecules make extensive hydrophobic contact with their apolar face stacking onto the flat surfaces of aromatic rings such as those of tryptophan and tyrosine residues. Such interactions between carbohydrate molecules and aromatic residues in proteins have been well documented (Bundle & Young, 1992; Spurlino *et al.*, 1992; Bourne *et al.*, 1993). Lawson *et al.* (1994) have shown that this type of interaction occurs in the first binding site with the planar side-chain rings of W616 and W662 stacked against separate glucose rings of the maltose molecule and similarly in the second binding site, the reducing sugar is stacked against the aromatic ring of Y633.

Additionally, there are numerous hydroxyl groups in maltose available for hydrogen bonding directly to the protein or indirectly through water molecules. In site 1, there are direct hydrogen bonds to side-chain nitrogen and oxygen atoms of K651 and N667, other protein-maltose interactions mediated through water molecules involving CGTase residues S382 (outside domain E), W616 and E663, as well as an intramolecular hydrogen bond between the reducing and non-reducing sugars of maltose. In site 2, T598, A599, G601, N627 and Q628 interact directly with maltose while N603 is hydrogen bonded to the protein through water molecules.

We have studied the possible binding residues in SBD by superimposing the solution structure (SBD_{av-min}) onto 1CDG and analysing distances between homologous residues and the maltose molecules. No energy minimisation or refinement was carried out prior to this analysis. In site 1 many similar interactions are plausible. Firstly the hydrophobic stacking interaction is likely to be significant since W543 and W590 of SBD are located near W616 and W662 of 1CDG at similar distance and orientation of the indole rings. Direct interactions between the side-chain nitrogen and oxygen atoms of N595 of SBD and the O2 and O3 atoms of maltose can be proposed based on the N667-maltose interaction reported in 1CDG. The backbone of K578 also occupies similar conformational space to K651 of 1CDG. However the interaction between its N^ε atom and O3' of maltose (distance of 3.2 Å) observed in the latter structure is not feasible in SBD (6.3 Å). This may be due to the flexibility of the lysine side-chain observed in solution. Closer inspection of the family of 53 structures shows that this side-chain exists in three different conformations. In the majority of structures the N^ε atom of K578 is in fact close enough to

interact with the O2 (instead of O3') group of maltose. There is a single interaction between the backbone carbonyl oxygen of E663 (1CDG) and the O2' of maltose which is mediated by a water molecule. In SBD, E591 would be the candidate for a similar interaction except that the carbonyl oxygen to O2' distance (2.8 Å) suggests that there may be a direct interaction between the protein and carbohydrate. In addition, the side-chain oxygen atoms of residue E591 are capable of forming strong interactions with O2 and O3' of maltose. This interaction in SBD is the only one where a residue can interact with both the reducing and non-reducing sugars of maltose. The latter may be a replacement interaction for K578 N^ε to O3' (maltose) which, as noted above, is not feasible in SBD. In site 2, although SBD residues T525, T526 and N530 are in the vicinity of T598, A599 and N603, respectively, of 1CDG their interactions with maltose may be weaker as they are further away. The distance between N530 N^{δ2} in SBD and O3 of maltose is 4.8 Å but this interaction may still be possible since the equivalent interaction in 1CDG is mediated by water molecules. The side-chains of N627 and Q628 of 1CDG are thought to interact with maltose, however, the side-chains of the homologous SBD residues D554 and K555, respectively, are pointing in the opposite direction. These residues being on the surface are still capable of forming at least a weak interaction, particularly with a larger substrate molecule (e.g. starch) which may be capable of forming additional interactions (compared to maltose). The amide group of G601 (1CDG) and G528 (SBD) can form almost identical interactions to O2 of maltose. In SBD this interaction is likely to be strengthened by the oxygen atom of G528 also forming a hydrogen bond to the O2 group of maltose, perhaps in place of a T526 carbonyl oxygen to O2 (maltose) interaction which is not possible in SBD although an equivalent interaction is reported for 1CDG. Residue E529 in SBD, through its backbone N and O atoms, is capable of forming relatively strong interactions with the O2 of maltose. Analogous interactions were not reported in 1CDG. Furthermore, the group of residues in binding site 2 of 1CDG does not include tryptophan. In SBD however, we believe W563 is a binding site residue based on (1) results from our titration study, (2) the residue is conserved in all *Aspergillus* subfamily glucoamylase sequences (Coutinho & Reilly, 1994a), and (3) the homologous tryptophan has been proposed to be a binding site residue in other CGTase structures (Kubota *et al.*, 1991; Coutinho & Reilly, 1994b). Finally in site 2, the aromatic ring of Y633 is seen to stack against one of the sugar rings of maltose but this is not possible in SBD where the equivalent residue is D560. It is feasible that Y527 forms a hydrophobic interaction with maltose in place of D560. The aromatic ring of this tyrosine located at the tip of a loop, is protruding away from the SBD molecule and unsurprisingly, can adopt two conformations. In the population of conformers

which allows this interaction, the aromatic ring can stack over the non-reducing glucose unit. This interaction is not as well defined as other stacking interactions predicted in site 1 (residues W543 and W590).

The above analysis has shown that the interactions involving site 1 are better conserved between SBD and 1CDG than those in site 2. Overall in SBD the interactions with maltose may be expected to be stronger in site 1 due to the greater number of hydrogen bonds plausible between the protein and carbohydrate, including one residue (E591) which bridges the two glucose rings. The hydrophobic interaction in this site is also of considerable significance involving two tryptophan residues. These observations are consistent with proposals that although both sites are involved in ligand binding, they have different roles or mechanisms. For example from kinetic studies of CGTase with β CD, site 1 has been implicated as being more important for its role in binding or anchoring the ligand while site 2 is involved in kinetic events in association with the catalytic site (Dijkhuizen *et al.*, 1995).

Conclusion

The solution structure of the SBD of G1 from *A. niger* has been determined by NMR spectroscopy and X-PLOR using a simulated annealing protocol. The structure consists of eight β -strands arranged mainly in antiparallel fashion forming an open-sided β -barrel. The features are consistent with our previous study (Jacks *et al.*, 1995) except that the third strand does not adopt a β -bulge conformation.

Two possible substrate binding sites have been identified on the basis of results obtained from ligand titration experiments and also modelling the SBD solution structure to the homologous domain of the crystal structure of the CGTase/maltose complex (Lawson *et al.*, 1994). The residues implicated in binding are, in site 1, W543, K578, W590, E591, N595 and in site 2, T526, Y527, G528, E529, N530, D554, W563. The molecule is better defined where the binding sites are located (closer to the C-terminal end). The region near the N-terminal end where SBD attaches to the linker in the intact enzyme shows a degree of flexibility.

The findings presented here will form the basis for the dynamics studies which are currently in progress. From this study and work on SBD mutants and SBD/ligand complexes, we hope to further our understanding on the structure-function relationship of this protein as well as the nature and mechanism of ligand binding and facilitation of hydrolysis.

Materials and Methods

Sample preparation

Unlabelled SBD was obtained by proteolytic cleavage of G1 (Sigma) followed by purification as previously

described (Le Gal-Coëffet *et al.*, 1995). Uniformly ^{15}N -labelled SBD was expressed in *A. niger* using a pIGF fusion vector. The fungus was grown in minimal medium using $^{15}\text{NH}_4\text{Cl}$ as the sole nitrogen source. The expression and purification protocol for the isotopically labelled SBD has been described in detail elsewhere (Le Gal-Coëffet *et al.*, 1995; MacKenzie *et al.*, 1996). Typically, samples for NMR experiments contained 1 to 2 mM protein in 90% H_2O /10% $^2\text{H}_2\text{O}$ or 99.9% $^2\text{H}_2\text{O}$ solution. The pH of the sample in H_2O was measured at ambient temperature to be 5.2.

NMR experiments and data processing

All spectra were recorded on a Bruker AMX 500 spectrometer at 313 K. Additional 2D spectra were recorded at 303 K to resolve ambiguities arising from overlap. Quadrature detection in the indirectly detected dimensions was obtained by phase cycling the appropriate pulses according to the time-proportional phase incrementation (TPPI; Marion & Wüthrich, 1983) or States-TPPI (Marion *et al.*, 1989a) method. The H_2O signal was suppressed by low-power presaturation during the recycle delay and the ^1H carrier was placed at the solvent frequency. ^1H chemical shifts were referenced to the H_2O resonance at 4.63 ppm (313 K) or 4.74 ppm (303 K) relative to sodium 3-trimethylsilyl-2,2,3,3-($^2\text{H}_4$)propionate. ^{15}N chemical shifts were referenced indirectly by using the above ^1H frequencies for the H_2O resonance and the gyromagnetic ratios (Wishart *et al.*, 1995).

Homonuclear and heteronuclear 2D/3D experiments on unlabelled and ^{15}N -labelled SBD samples were recorded and processed as described previously (Jacks *et al.*, 1995). In addition, a primitive exclusive correlation spectroscopy (P.E.COSY; Mueller, 1987) spectrum was recorded with 530 (real) t_1 increments of 4096 (complex) data points using the ^{15}N -labelled sample which had been exchanged in $^2\text{H}_2\text{O}$. Coupling constants between the backbone nitrogen and H^β ($^3J_{\text{N}\beta}$) were estimated using the HNHB experiment (Archer *et al.*, 1991; Madsen *et al.*, 1993) which was acquired with constant time during the t_1 (^{15}N) evolution period. The data were acquired as 128 (complex), 64 (real) and 512 (real) points with spectral widths of 6250 (^1H), 1667 (^{15}N) and 6250 (^1H) Hz in t_1 , t_2 and t_3 , respectively.

NMR data were processed on Silicon Graphics workstations using the FELIX software package (*Felix User Guide*, version 2.3, Biosym Technologies, San Diego, 1993). Removal of the low-frequency component of the time-domain data (Marion *et al.*, 1989b; Waltho & Cavanagh, 1993) was performed prior to apodisation and Fourier transformation in order to remove the residual H_2O signal. The P.E.COSY was processed with an 80° or 90° -shifted sine-bell function in both dimensions to yield a final matrix of 4096 (F_2) $\times$ 2048 (F_1) real points. The HNHB data were processed with a shifted sine-bell function in all dimensions resulting in a matrix of 512 (H^N , F_3) $\times$ 64 (^{15}N , F_2) $\times$ 512 (^1H , F_1) real points. The first point in each experiment was multiplied by 0.5 before Fourier transformation to suppress ridges in the transformed spectrum (Otting *et al.*, 1986). In F_3 , only the downfield half (amide and aromatic region) of the spectral width was retained as the data points upfield of the H_2O signal were discarded after Fourier transformation.

The effect of titrations with β CD was monitored by one-dimensional (1D) ^1H and 2D clean total correlation spectroscopy (TOCSY; Griesinger *et al.*, 1988) spectra

using the unlabelled SBD sample and by 2D ^1H - ^{15}N HSQC spectra (Bodenhausen & Ruben, 1980) of uniformly ^{15}N -labelled SBD. Titrations with MH were followed by 1D ^1H spectra only using the unlabelled protein sample. Experiments were performed at 300 or 310 K. 1D spectra were typically recorded with 256 scans and 16,384 complex points over a spectral width of 12,500 Hz. The 2D spectra were recorded as described previously (Jacks *et al.*, 1995) except that for TOCSY experiments, 256 to 512 real t_1 increments and a mixing time of 60 ms were used and for HSQC spectra, 200 to 256 real t_1 increments were recorded. Spectra were acquired successively for each ligand titration series. The time-domain data of TOCSY experiments were apodised by a 40° -shifted sine-bell or sine-squared function, followed by zero-filling, to obtain 2048×2048 real points. The HSQC data were processed with a 60° -shifted sine-squared function and the final matrix size was 1024×1024 real points.

Distance constraints

Distance constraints were derived manually from a 2D NOE spectroscopy (NOESY) spectrum acquired with a mixing time of 90 ms. Upper bounds were calibrated by counting the number of contour levels of non-overlapped cross-peaks arising from protons in regions of well defined secondary structure (as determined by Jacks *et al.*, 1995) and correlating these with known distances. Sequential, parallel cross-strand and antiparallel cross-strand interactions were considered. Typically, eight to ten cross-peaks were used to obtain the average intensity (number of contour levels) and interproton interactions in different parts of the molecule were selected wherever possible. This resulted in the following upper bounds: ≥ 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 115 intraresidue ($i = j$), 290 sequential ($|i - j| = 1$), 100 medium-range ($2 \leq |i - j| \leq 4$), and 587 long-range ($|i - j| \geq 5$) NOEs.

Stereospecific assignment and dihedral angle constraints

Vicinal coupling constants, $^3J_{\text{N}\alpha}$, were measured from double quantum filtered correlation spectroscopy (DQF-COSY) and NOESY spectra using the PRONTO/3D software (Pronto Software Development and Distribution, Copenhagen, Denmark; Kjær *et al.*, 1994) in which J values are determined by the method of Ludvigsen *et al.* (1991). Where a realistic value could not be obtained due to weak or absent cross-peaks from one or both spectra, the line fitting interface in FELIX was used. Simulated lineshapes were fitted to F_2 cross-sections (which had been zero-filled to 32,768 points) of NOESY and TOCSY cross-peaks using the simulated annealing optimisation method and the results from the two spectra were compared for consistency wherever possible. In the structure calculations, for residues with $^3J_{\text{N}\alpha}$ values ≥ 8.0 Hz, < 8.0 Hz but ≥ 7.0 Hz, and ≤ 5.5 Hz, the corresponding ϕ angles were constrained to $-120^\circ \pm 40^\circ$, $-120^\circ \pm 50^\circ$ and $-65^\circ \pm 25^\circ$, respectively. For residues which occur in the middle of β -strands and have small (≤ 5.5 Hz) $^3J_{\text{N}\alpha}$ values the ϕ constraints were included and their feasibility checked in the resulting structures.

Stereospecific assignments of prochiral β -methylene groups were obtained by identifying predominant χ_1

rotamers wherever possible. This was based on H^α - H^β coupling constants ($^3J_{\alpha\beta}$) measured from the P.E.COSY, the relative sizes of $^3J_{\text{N}\beta}$ values from the HNHB spectrum, and the relative intensities of H^α - H^β and H^N - H^β NOEs for each β -methylene proton pair. In some cases, where J values could not be measured from the P.E.COSY, the relative intensities of H^α - H^β cross-peaks were compared in the DQF-COSY spectrum. The χ_1 restraints were set to $\pm 30^\circ$ of the appropriate staggered (-60° , 60° , 180°) rotamer. When the H^α - H^β peak and/or the H^N - H^β peak used in the stereospecific assignment was overlapped, a tolerance of $\pm 40^\circ$ was allowed on the χ_1 restraint. In cases where only sufficient data were available to exclude the 60° rotamer, χ_1 was restrained to $-120^\circ \pm 90^\circ$ without stereospecifically assigning the β -methylene protons. Of the residues in this category, some converged to a single rotamer conformation in subsequent calculations; these were then stereospecifically assigned and the appropriate χ_1 restraint applied after checking the feasibility of intra and interresidue NOEs affected by this restraint.

For valine residues, stereospecific assignments of the γ -methyl groups and χ_1 angle constraints were derived from $^3J_{\alpha\beta}$ values and intensities of pairs of H^α - H^γ and H^N - H^γ NOEs. The χ_1 angle was constrained to $\pm 30^\circ$ of the appropriate staggered rotamer. Similar χ_1 angle constraints were applied to threonine residues.

For leucine residues, stereospecific assignment of the δ -methyl groups was obtained if the χ_2 angle was well defined following preliminary calculations with the χ_1 angle constrained. When the χ_2 angle converged to a single rotamer in 80 to 100% of the preliminary structures, using a combination of H^N - H^δ , H^α - H^δ and H^β - H^δ NOEs, the χ_2 constraint was assigned to this rotamer ($\pm 40^\circ$) in subsequent calculations. Structures calculated were analysed to confirm that the χ_2 constraint was satisfied and that significant violations were not introduced.

Hydrogen bonds

In initial rounds of calculations, hydrogen bond constraints were not included. Preliminary calculated structures were examined for possible hydrogen bond donor-acceptor pairs. When such pairs were consistent with patterns inferred from amide exchange and NOE data (Jacks *et al.*, 1995), they were included in subsequent calculations as distance constraints of $d(\text{H}-\text{O}) = 1.8$ to 2.5 Å and $d(\text{N}-\text{O}) = 2.5$ to 3.3 Å resulting in two constraints per hydrogen bond. Previously unidentified long-range pairs were also incorporated at this stage if they were present in 80 to 100% of the structures and were supported by amide exchange and NOE data. Structures from subsequent calculations were analysed to ensure that no large violations resulted from incorporating these constraints.

Structure calculations

Structure calculations were performed with a simulated annealing protocol using the X-PLOR program (Brünger, 1992) on a Silicon Graphics workstation. The *topallhdg.pro* topology file and *parallhdg.pro* parameter file within the program were used. The disulphide bridge between C509 and C604 was incorporated during structure generation using the default distance value. A time step of 1 fs was used throughout the calculation. The protocol commenced with generation of starting structures with random coordinates. For this stage, the

soft square well potential was used for distance constraints and force constants were set to very low values. For the simulated annealing stage, the square well potential was used for distance constraints. Initially 100 steps of Powell minimisation were performed with only bond, van der Waals and distance restraint energy terms. An additional 100 steps of minimisation were performed including the angle energy term. Chirality and planarity terms were then introduced and 3 ps of dynamics performed at 2000 K during which time the force constants were gradually increased except for the repulsive force constant which was decreased to a final value of $0.003 \text{ kcal mol}^{-1} \text{ \AA}^{-4}$. A period of cooling to 100 K followed over 4.94 ps (38 steps of 0.13 ps with 50 K cooling/step) where the repulsive force constant was gradually increased to a final value of $4.0 \text{ kcal mol}^{-1} \text{ \AA}^{-4}$. Powell minimisation was then carried out for 200 steps. For the refinement stage the square well potential was used once again. Slow cooling was implemented (1500 K to 100 K over 28 steps of 0.89 ps dynamics with 50 K cooling/step) to improve refinement. Finally, 1000 steps of Powell minimisation were performed. The sum averaging option was used to treat equivalent and non-stereospecifically assigned protons.

Calculations were conducted iteratively by checking violations after each round and adding new NOE constraints. When most cross-peaks had been assigned unambiguously as far as possible, the average structure calculated from a family of ten structures was used to extract all interproton distances less than 5 Å. At this stage the global fold was determined with the family of ten structures showing a mean rmsd to the average structure (calculated over the secondary structure elements) of approximately 1 Å. From the extracted distances, further, weak NOEs were assigned and included as distance constraints in subsequent calculations. New NOEs assigned in this manner were used conservatively in order to avoid introducing a bias towards the average structure. For the final round of calculations, 100 structures were calculated and refined. The resulting structures were ranked in order of rmsd of NOE constraints to select structures for further analysis.

Analysis of structures

Graphical manipulation and analysis of the family of calculated structures were performed on a Silicon Graphics workstation using the *Insight II*[®] molecular modelling system.

Structures were analysed and selected on the basis of number of distance and dihedral constraint violations and low energies. From the family of 53 refined structures of SBD, a structure of mean coordinates was calculated using X-PLOR; this is referred to as SBD_{av}. This structure was also energy-minimised by 1500 steps of restrained Powell minimisation to obtain SBD_{av-min}. The coordinates of five final structures selected at random (1KUL), coordinates of SBD_{av-min} (1KUM) and the list of NMR constraints (R1KULMR) have been deposited in the Brookhaven Protein Data Bank.

For structural comparisons, individual structures from the ensemble were superimposed onto SBD_{av}. Comparisons were also made between SBD_{av-min} and two crystal structures (Klein & Schulz, 1991; Lawson *et al.*, 1994) of the homologous domain of CGTase.

Dihedral angle order parameters (S) for ϕ , ψ and χ_1 angles were calculated according to Hyberts *et al.* (1992). A well defined dihedral angle is reflected by an S value approaching unity.

The solvent accessible area for SBD_{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.

Ligand-binding experiments

(+)-1-Deoxynojirimycin hydrochloride, βCD, MH and ²H₂O were purchased from Sigma Chemical Company.

Unlabelled SBD samples of 1.0 or 2.2 mM for NMR were freshly prepared by dissolving in 90% H₂O/10% ²H₂O. The sample used for the MH titration was 1.0 mM SBD and also contained 1.0 mM (+)-1-deoxynojirimycin in order to minimise the degradation of MH by any residual G1 which may be present in the SBD sample. Stock solutions of βCD (14.1 or 16.3 mM) and MH (28.6 mM) were prepared in 100% H₂O from which small aliquots were titrated into the SBD sample. The addition of ligand was carried out, initially, in small steps of 0.1 to 0.2 molar equivalence relative to the SBD sample, increasing to steps of 0.2 to 0.5 molar equivalence until a ratio of 5:1 or 2:1 (ligand:SBD) was reached for 1D and 2D experiments, respectively.

Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC; LINK grant no. LR50/587 and grant no. 50/C673-I).

References

- Aleshin, A., Golubev, A., Firsov, L. M. & Honzatko, R. B. (1992). Crystal structure of glucoamylase from *Aspergillus awamori* var. *X100* to 2.2-Å resolution. *J. Biol. Chem.* **267**, 19291–19298.
- Aleshin, A. E., Hoffman, C., Firsov, L. M. & Honzatko, R. B. (1994a). Refined crystal structures of glucoamylase from *Aspergillus awamori* var. *X100*. *J. Mol. Biol.* **238**, 575–591.
- Aleshin, A. E., Firsov, L. M. & Honzatko, R. B. (1994b). Refined structure for the complex of acarbose with glucoamylase from *Aspergillus awamori* var. *X100* to 2.4-Å resolution. *J. Biol. Chem.* **269**, 15631–15639.
- Archer, S. J., Ikura, M., Torchia, D. A. & Bax, A. (1991). An alternative 3D NMR technique for correlating backbone ¹⁵N with side-chain Hβ resonances in larger proteins. *J. Magn. Reson.* **95**, 636–641.
- Bahl, M., Burchhardt, G., Spreinat, A., Haeckel, K., Weinecke, A., Schmidt, B. & 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.
- Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters*, **69**, 185–189.
- Bourne, Y., van Tilbeurgh, H. & Cambillau, C. (1993). Protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* **3**, 681–686.
- Brünger, A. T. (1992). *X-PLOR Version 3.1: A system for Crystallography and NMR*, Yale University, New Haven.
- Bundle, D. R. & Young, N. M. (1992). Carbohydrate-protein interactions in antibodies and lectins. *Curr. Opin. Struct. Biol.* **2**, 666–673.

- Clarke, A. J. & Svensson, B. (1984). The role of tryptophanyl residues in the function of *Aspergillus niger* glucoamylase G1 and G2. *Carlsberg Res. Commun.* **49**, 111–122.
- Coutinho, P. M. & Reilly, P. J. (1994a). Structure-function relationships in the catalytic and starch binding domains of glucoamylase. *Protein Eng.* **7**, 393–400.
- Coutinho, P. M. & Reilly, P. J. (1994b). Structural similarities in glucoamylases by hydrophobic cluster analysis. *Protein Eng.* **7**, 749–760.
- 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 & Complementary Technologies* (Geisow, M. J. & Epton, R., eds), pp. 96–99, Mayflower Worldwide Limited, Birmingham.
- Din, N., Forsythe, I. J., Burtnick, L. D., Gilkes, N. R., Miller, R. C., Jr, Warren, R. A. J. & Kilburn, D. G. (1994). The cellulose-binding domain of endoglucanase A (CenA) from *Cellulomonas fimi*: evidence for the involvement of tryptophan residues in binding. *Mol. Microbiol.* **11**, 747–755.
- Gilkes, N. R., Warren, R. A. J., Miller, R. C., Jr & Kilburn, D. G. (1988). Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. *J. Biol. Chem.* **263**, 10401–10407.
- Goto, M., Semimaru, T., Furukawa, K. & Hayashida, S. (1994). Analysis of the raw starch-binding domain by mutation of a glucoamylase from *Aspergillus awamori* var. *kawachi* expressed in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **60**, 3926–3930.
- Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R. R. (1988). Clean TOCSY for ^1H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110**, 7870–7872.
- Harris, E. M. S., Aleshin, A. E., Firsov, L. M. & Honzatko, R. B. (1993). Refined structure for the complex of 1-deoxynojirimycin with glucoamylase from *Aspergillus awamori* var. X100 to 2.4-Å resolution. *Biochemistry*, **32**, 1618–1626.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Klein, C. & Schulz, G. E. (1991). Structure of cyclodextrin glycosyltransferase refined at 2.0 Å resolution. *J. Mol. Biol.* **217**, 737–750.
- Kramer, G. F. H., Gunning, A. P., Morris, V. J., Belshaw, N. J. & Williamson, G. (1993). Scanning tunneling microscopy of *Aspergillus niger* glucoamylases. *J. Chem. Soc., Faraday Trans.* **89**, 2595–2602.
- Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
- Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J. & Gronenborn, A. M. (1989). Determination of the three-dimensional structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. *Biochemistry*, **28**, 7241–7257.
- Kubota, M., Matsuura, Y., Sakai, S. & Katsube, Y. (1991). Molecular structure of *B. stearotheophilus* cyclodextrin glucanotransferase and analysis of substrate binding site. *Denpun Kagaku*, **38**, 141–146.
- Kusnadi, A. R., Chang, H. Y., Nikolov, Z. L., Metzler, D. E. & Metzler, C. M. (1994). Starch-binding domain of *Aspergillus* glucoamylase-I. Interaction with β -cyclodextrin and maltoheptaose. *Ann. N.Y. Acad. Sci.* **721**, 168–177.
- 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. Crystallog.* **26**, 283–291.
- Lawson, C. L., van Montfort, R., Strokopytov, B., Rozeboom, H. J., Kalk, K. H., de Vries, G. E., Penninga, D., Dijkhuizen, L. & 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.
- 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.
- Ludvigsen, S., Andersen, K. V. & Poulsen, F. M. (1991). Accurate measurements of coupling constants from two-dimensional nuclear magnetic resonance spectra of proteins and determination of ϕ -angles. *J. Mol. Biol.* **217**, 731–736.
- Madsen, J. C., Sørensen, O. W., Sørensen, P. & Poulsen, F. M. (1993). Improved pulse sequences for measuring coupling constants in ^{13}C , ^{15}N -labeled proteins. *J. Biomol. NMR*, **3**, 239–244.
- 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.* In the press.
- Marion, D. & Wüthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^1H - ^1H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**, 967–974.
- Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989a). Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393–399.

- Marion, D., Ikura, M. & Bax, A. (1989b). Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data. *J. Magn. Reson.* **84**, 425–430.
- Mueller, L. (1987). P.E.COSY, a simple alternative to E.COSY. *J. Magn. Reson.* **72**, 191–196.
- 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.
- 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.
- 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.
- 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.
- Spurlino, J. C., Rodseth, L. E. & Quijcho, F. A. (1992). Atomic interactions in protein-carbohydrate complexes. Tryptophan residues in the periplasmic maltodextrin receptor for active transport and chemotaxis. *J. Mol. Biol.* **226**, 15–22.
- Ståhlberg, J., Johansson, G. & Pettersson, G. (1991). A new model for enzymatic hydrolysis of cellulose based on the two-domain structure of cellobiohydrolase I. *Bio/Technology*, **9**, 286–290.
- Stoffer, B., Aleshin, A. E., Firsov, L. M., Svensson, B. & Honzatko, R. B. (1995). Refined structure for the complex of D-glucosyl-dihydroacarbonyl with glucoamylase from *Aspergillus awamori* var. X100 to 2.2 Å resolution: dual conformations for extended inhibitors bound to the active site of glucoamylase. *FEBS Letters*, **358**, 57–61.
- Strokopytov, B., Penninga, D., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L. & Dijkstra, B. W. (1995). X-ray structure of cyclodextrin glycosyltransferase complexed with acarbose. Implications for the catalytic mechanism of glycosidases. *Biochemistry*, **34**, 2234–2240.
- Svensson, B. (1988). Regional distant sequence homology between amylases, α -glucosidases and transglucanases. *FEBS Letters*, **230**, 72–76.
- 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.
- 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.
- Svensson, B., Jespersen, H., Sierks, M. R. & MacGregor, E. A. (1989). Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. *Biochem. J.* **264**, 309–311.
- 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. (Tokyo)*, **98**, 663–671.
- Waltho, J. P. & Cavanagh, J. (1993). Practical aspects of recording multidimensional NMR spectra in water with flat baselines. *J. Magn. Reson. A*, **103**, 338–348.
- Watanabe, T., Oyanagi, W., Suzuki, K. & Tanaka, H. (1990). Chitinase system of *Bacillus circulans* WL-12 and importance of chitinase A1 in chitin degradation. *J. Bacteriol.* **172**, 4017–4022.
- 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.
- Williamson, M. P. & Asakura, T. (1993). Empirical comparisons of models for chemical shift calculations in proteins. *J. Magn. Reson. B*, **101**, 63–71.
- Williamson, G., Belshaw, N. J. & Williamson, M. P. (1992). O-Glycosylation in *Aspergillus* glucoamylase. Conformation and role in binding. *Biochem. J.* **282**, 423–428.
- Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995). ^1H , ^{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.
- Xu, G.-Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E. & Harvey, T. S. (1995). Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry*, **34**, 6993–7009.

Edited by P. E. Wright

(Received 29 January 1996; received in revised form 1 April 1996; accepted 22 April 1996)