J. Mol. Biol. (1998) **275**, 491–502

JMB

Crystal Structure of β-Glucosidase A from *Bacillus polymyxa*: Insights into the Catalytic Activity in Family 1 Glycosyl Hydrolases

J. Sanz-Aparicio[1]*, J. A. Hermoso[1], M. Martínez-Ripoll[1], J. L. Lequerica[2] and J. Polaina[2]

[1]*Grupo de Cristalografía Macromolecular y Biología Estructural, Instituto de Química-Física ''Rocasolano'' CSIC, Serrano 119 28006, Madrid, Spain*

[2]*Instituto de Agroquímica y Tecnología de Alimentos, CSIC Apdo. 73, 46100-Burjassot Valencia, Spain*

**Corresponding author*

Family 1 glycosyl hydrolases are a very relevant group of enzymes because of the diversity of biological roles in which they are involved, and their generalized occurrence in all sorts of living organisms. The biological plasticity of these enzymes is a consequence of the variety of β-glycosidic substrates that they can hydrolyze: disaccharides such as cellobiose and lactose, phosphorylated disaccharides, cyanogenic glycosides, etc. The crystal structure of BglA, a member of the family, has been determined in the native state and complexed with gluconate ligand, at 2.4 Å and 2.3 Å resolution, respectively. The subunits of the octameric enzyme display the $(\alpha/\beta)_8$ barrel structural fold previously reported for other family 1 enzymes. However, significant structural differences have been encountered in the loops surrounding the active-center cavity. These differences make a wide and extended cavity in BglA, which seems to be able to accommodate substrates longer than cellobiose, its natural substrate. Furthermore, a third sub-site is encountered, which might have some connection with the *trans*glycosylating activity associated to this enzyme and its certain activity against β-1,4 oligosaccharides composed of more than two units of glucose. The particular geometry of the cavity which contains the active center of BglA must therefore account for both, hydrolytic and *trans*glycosylating activities. A potent and well known inhibitor of different glycosidases, D-glucono-1,5-lactone, was used in an attempt to define interactions of the substrate with specific protein residues. Although the lactone has transformed into gluconate under crystallizing conditions, the open species still binds the enzyme, the conformation of its chain mimicking the true inhibitor. From the analysis of the enzyme-ligand hydrogen bonding interactions, a detailed picture of the active center can be drawn, for a family 1 enzyme. In this way, Gln20, His121, Tyr296, Glu405 and Trp406 are identified as determinant residues in the recognition of the substrate. In particular, two bidentate hydrogen bonds made by Gln20 and Glu405, could conform the structural explanation for the ability of most members of the family for displaying both, glucosidase and galactosidase activity.

Keywords: family 1 glycosyl hydrolase; enzyme-ligand complex; substrate recognition; X-ray structure

Introduction

The name β-glucosidase is given to different types of enzymes capable of hydrolyzing β-glucosidic linkages of disaccharides, oligosaccharides or conjugated glucosides. According to a classification of glycosyl hydrolases based on their amino acid sequence similarities, which takes into account evolutionary relationships, β-glucosidases appear in five of the 57 families whose existence has been ascertained in recent surveys (Henrissat, 1991; Henrissat & Bairoch, 1993, 1996). Differences in substrate specificity and evolutionary origin reflect the different biological roles played by these

0022–2836/98/030491–12 $25.00/0/mb971467

enzymes. A large number of β-glucosidases, together with other β-glycosidases, are grouped in family 1 of Henrissat's classification. Outstanding characteristics of enzymes belonging to this important family are a wide range of substrate specificity and a quite high degree of sequence homology, which is a remarkable fact considering that these enzymes are present in all kinds of living organisms: bacteria, archaea, plants and animals. The hydrolysis of the β-glycosydic linkage is carried out with net retention of the anomeric configuration, likely *via* a double displacement mechanism involving the formation of a covalent glycosyl-enzyme intermediate (Kempton & Withers, 1992; Sinnot, 1990; Wang *et al.*, 1995; Withers *et al.*, 1992).

On the basis of similarities at the level of tertiary structure, families of glycosyl hydrolases are classified in groups designated superfamilies or clans. Family 1 is included in clan GH-A (also named superfamily 4/7) characterized by an 8-fold α/β barrel structure, in which the two amino acids of the active site directly involved in the catalysis, the acid/base catalyst and the nucleophile, are located in β strands number 4 and 7, respectively (Henrissat *et al.*, 1995; Henrissat & Bairoch, 1996; Henrissat & Romeu, 1995; Jenkins *et al.*, 1995) The three-dimensional structures of three family 1 enzymes have been reported so far: a phospho-β-galactosidase from *Lactococcus lactis* (Wiesmann *et al.*, 1995), a cyanogenic β-glucosidase from *Trifolium repens* (Barrett *et al.*, 1995), and very recently, a myrosinase from *Sinapsis alba* (Burmeister *et al.*, 1997). The first enzyme (henceforth referred as 1PBG) present in lactic and other homofermentative bacteria, is responsible for the hydrolysis of phosphorylated lactose which enters the cell by a phosphotransferase system. The second enzyme (henceforth 1CBG) is involved in cyanogenesis, a defense mechanism present in many plant species to avoid attack by pathogens or grazing by herbivors. The third enzyme (henceforth 1MYR), is characteristic of plant species of the family *Cruciferae*. Its biological role seems to be similar to the enzyme from *Trifolium*, as it hydrolyzes thio-β-glycosides to yield toxic compounds. The tertiary structure of these enzymes confirmed the prediction of an α/β barrel made for members of the GH-A clan, and allowed the identification of the catalytic residues. However, the elucidation of the chemical basis accounting for the differences in substrate specificity and affinity shown by family 1 enzymes is not fully understood and requires the identification of the residues involved in specific enzyme-substrate interactions.

This paper presents the structure of β-glucosidase A from *Bacillus polymyxa* (henceforth BglA) and a comparative analysis with the other members of the same family whose coordinates are available. The complex of the enzyme with a ligand and a detailed analysis of enzyme-substrate interactions are shown. BglA represents another subgroup of family 1 enzymes present in many bacterial species involved in the hydrolysis of cellulosic substrates. Preliminary crystallographic data of this enzyme have been reported (Sanz-Aparicio *et al.*, 1994). Two genes from *B. polymyxa* encoding homologous β-glucosidases BglA and BglB have been isolated. The analysis of their sequences showed the existence of the group of enzymes later labeled family 1 in Henrisat's classification (González-Candelas *et al.*, 1990). Biochemical characterization of BglA showed that the native protein is an intracellular enzyme with an octameric configuration and a molecular mass of about 400,000 Da (Sanz-Aparicio *et al.*, 1994). As it occurs with other glycosidases, the enzyme catalyzes *trans*glycosylation to sugar acceptors (Painbeni *et al.*, 1992). The expression of BglA activity can be easily assayed in *Escherichia coli* by using chromogenic substrates. This property makes BglA a convenient reporter enzyme and a suitable system to study problems related to protein engineering: it allows the designing of experiments in which thousands of BglA-expressing colonies can be screened in search of mutants generated by random or site-directed mutagenesis (López-Camacho *et al.*, 1996).

Results and Discussion

Quality of the model

The crystal structure of β-glucosidase from *B. polymyxa* has been determined by the molecular replacement method (see Materials and Methods) and has been refined against diffraction data up to 2.4 Å resolution (Table 1). The crystals belong to the space group $P4_22_12$, with strong pseudo-$C222_1$ symmetry. Diffraction patterns collected using the weaker radiation from a rotating-anode generator could only be indexed in a centered-cell, as it was reported earlier (Sanz-Aparicio *et al.*, 1994). True symmetry did not become apparent until synchrotron data were available. The crystals present four independent molecules in the asymmetric unit and 70% solvent content in the cell. The enzyme exists as an octamer in solution. In the crystal this octamer follows 422 point symmetry, with only one of the 2-fold axes being crystallographic and the other two axes being local. The final model contains residues 2 to 448 of each independent subunit (the first residue, Met1, is not present in the mature protein), 1541 water molecules and one phosphoryl group located at Cys169 of each chain. The model has good stereochemistry (Table 1) and all residues fall within the allowed regions of the Ramachandran plot. As the model was refined imposing non-crystallographic symmetry restraints between the four molecules of the asymmetric unit, the r.m.s. differences between coordinates after superposition of the four subunits was lower than 0.2 Å and, thus, all following discussion will be based on a single molecule.

Table 1. Data collection and refinement statistics

	Native BglA		BglA-gluconate
A. *Data collection*			
Temperature (K)	278	100	100
Source	Rotating-anode	Synchrotron	Synchrotron
Space group	$C222_1$[c]	$P4_22_12$	$P4_22_12$
Cell *a*, *b*, *c* (Å)	155.73, 208.40, 208.50	205.46,155.86	205.85,155.50
Resolution range (Å)	32.3–3.0	26.5–2.4	31.4–2.3
No. of crystals	4	1	1
No. of observations	16,8130	584,536	840,221
No. of unique refl.	57,489	117,344	146,539
Completeness of data (%)			
all data	85.1	97.3	99.4
outer shell	82.2	92.3	97.2
R_{sym}[a]	0.082	0.083	0.092
B. *Refinement statistics*			
Sigma cutoff	–	2	2
Resolution range (Å)	8.0–3.0	8.0–2.4	8.0–2.3
No. of protein atoms	14,576	14,576	14,576
No. of solvent atoms		1541	1494
R factor[b]	0.27	0.20	0.20
Free *R* factor	0.31	0.27	0.25
r.m.s. deviation			
bond lengths (Å)		0.009	0.009
bond angles (°)		1.51	1.56
dihedrals		23.99	24.01
impropers		1.35	1.38
bonded *B*-factors		3.00	3.06
Ramachandran outliers		None	None

[a] $R_{sym} = \Sigma(I - \langle I \rangle)/\Sigma(I)$.
[b] R-factor $= \Sigma||F_o| - |F_c||/\Sigma|F_o|$.
[c] Pseudosymmetry (see the text).

The monomer subunit and structural comparison with other family 1 glycosyl hydrolases

The monomeric subunit adopts the expected topology of a single $(\alpha/\beta)_8$ barrel, with additional units of secondary structure inserted between the α/β units (Figure 1). These extra regions of secondary structure are quite well conserved in the family 1 enzymes whose structure is known up to the present. In the next analysis, coordinates for 1PBG and 1CBG, the only two family 1 members available in the PDB, will be used for comparison with BglA. According to Burmeister *et al.* (1997), the recently reported structure of myrosinase (1MYR) is most similar to the enzyme from white clover (1CBG), both isolated from plants. Structural superposition of the three chains reveals that there are four loop regions that differ strongly in all three cases not only in the number of insertions detected (Figure 2) but also in their structural arrangements (Figures 3 and 4). These regions correspond to the long loops connecting barrel strands and helices (at the C terminus of the chain) and are next to the cavity defined by the center of the barrel. Apart from these loops, the overall folding and even loops located in the opposite face of the barrel are markedly well conserved. 1CBG has a long insertion of more than ten residues at the N terminus, its chain spanning next to the C terminus of the two bacterial enzymes.

The loop A (see Figures 3 and 4) is different in 1PBG with respect to the other two enzymes. This loop is situated over the segment defined as the site of the substrate phosphoryl moiety in the phosphoenzymes of the family (Wiesmann *et al.*, 1995). In 1PBG, the chain of the loop A is retracted in a way such that a polar region is created by three residues, Tyr37, Tyr42 and Tyr44, which are pointing to the entrance of the pocket. The environment at this point in BglA and 1CBG is, on the contrary, non-polar. A second region showing differences is loop B (Figures 3 and 4). There is an insertion of 12 residues in 1CBG, which makes a knot-like structure closed by the di-sulpur bridge Cys202–Cys210. The cavity of the cyanogenic enzyme looks somehow deeper because of the effect produced by this extra loop.

Larger differences between the three enzymes are found in the long loop, C, connecting strand 6 with helix 6 of the barrel (Figures 3 and 4). This loop forms a part of the cavity and must play an important role in the binding of the substrate. Therefore, these differences must account for the diversity of functions characteristic of family 1 enzymes. Mobility seems to be another feature of this region as suggested by two missing zones that could not be modeled from density in native 1PBG (Wiesmann *et al.*, 1995). The region Phe303 to Ser312 is much shorter in BglA and occupies an intermediate position between the extended open

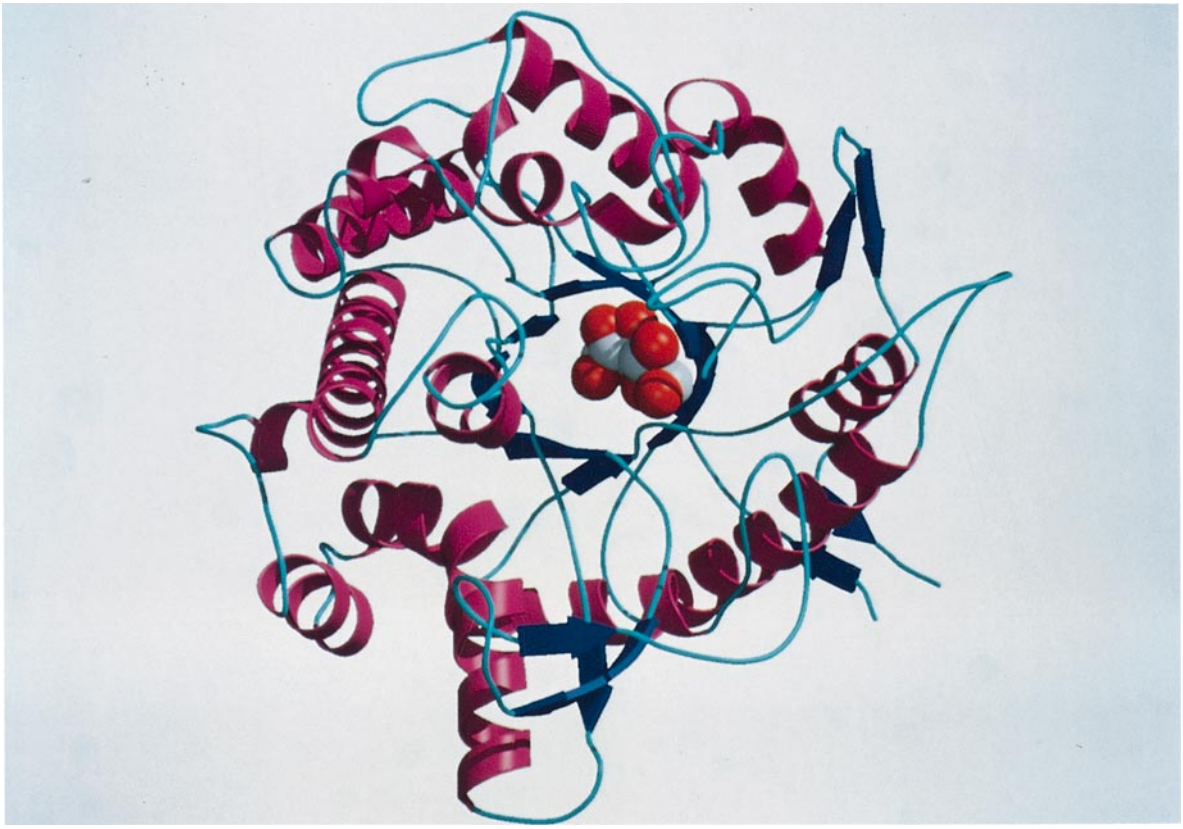

Figure 1. The overall folding of the subunit of the octameric β-glucosidase (BglA) from *Bacillus polymyxa*, complexed with a ligand. Ribbon representation of the main-chain trace, with helices shown in pink, β-strands in blue and loops in cyan. Gluconate moiety of the enzyme-ligand complex is displayed as CPK model, carbon atoms in white spheres and oxygens in red. Figure drawn using the program MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt & Murphy, 1994).

conformation of 1CBG and the long region of 1PBG that seems to close the access to the cavity. The end of the loop must be in close contact with the non-reducing end of the bound disaccharide (the hydroxynitrile group in 1CBG). In particular, Ser368 to Ile372 of 1CBG is protruding into the aglycon site, with Trp369 and one histidine of the strand 5 (His256) spanning part of this cavity which must accommodate a moiety smaller than in the case of both bacterial enzymes. Finally, the different conformation of the loop D in 1CBG (Figures 3 and 4) is possibly due to this region being involved in contacts between both subunits of the dimer in the cyanogenic enzyme, as this region is far from the active center.

As a consequence of the different structural arrangements adopted by the loops surrounding the active center, the cavity of BglA is wider and more extended than the one found in the other two enzymes and so it is possibly able to accommodate larger substrates. Starting from the structure of the enzyme-gluconate complex (see later), we have modeled the possible site for a second glucose ring (results not shown). In 1PBG, this second sugar ring would cover the cavity, while in BglA and 1CBG, a third pocket is encountered. This third sub-site is more open and extended in BglA, in which it is only limited by loop C. This fact might have some connection with BglA, despite being a cellobiase, having also some activity against β-1, 4-oligosaccharides composed of more than two units of glucose. On the other hand, an important property of this enzyme is *trans*glycosylation. In bacteria and fungi, the *trans*glycosylating activity associated to some glycosidases generates a signal which activates the transcription of genes controlling other enzymes (Beguin & Aubert, 1994; Painbeni *et al.*, 1992). For instance β-glucosidase from *Trichoderma reesei* is reported to be capable of forming sophorose by *trans*glycosylation. In turn, sophorose acts as an inducer of the cellulolytic system of the fungus. The geometry of the active site cavity of BglA, therefore, must reflect both, hydrolytic and *trans*glycosylating activities.

Quaternary structure

BglA presents an octameric aggregation state in solution but no cooperativity has been detected between the catalytic centers. The overall dimensions of the octamer are approximately 130 Å × 130 Å × 100 Å. Figure 5(a) shows the octamer found in the crystal as seen along the local 4-fold axis. A rotation of 90° (Figure 5(b)) reveals that closer interactions exist between subunits related by 2-fold symmetry generating pseudo-

Figure 2. Structurally different regions in family 1 enzymes. Alignment of the sequence in the non-conserved loops.

dimers. As a consequence, the octamer could be described as a tetramer of dimers oriented along the local 4-fold axis.

Although the orientation of the subunits in the dimer is very different in BglA with respect to 1CBG (Barrett *et al.*, 1995) the region of the molecule involved is roughly the same, but the interaction is much stronger in the BglA dimer, being maintained by a series of 14 hydrogen bonds. Helix 7 of the barrel and loop 228 to 232 make ten of these bridges and are mainly responsible for the interface interaction in the BglA dimer. His383, Arg384 and Asp388 of helix 7 contribute through their side-chains, while the loop residues Pro228, Tyr229 and Thr231 contribute with both side and main-chain atoms. The interface is also formed by the interaction between the loop 315 to 320 and residues in the N and C terminus of the chain, the main-chain of Asn316 making hydrogen bonds to Thr2 and Thr446 hydroxyl groups. There are also eight water molecules linking both subunits. The interface between units related by the 4-fold axis, on the contrary, is much less extended and occurs only at two points. The loop 47 to 50 interacts with the loop 441 to 447 through three hydrogen bonds, and Arg422 side-chain links to Lys365 and Gln430. There are also five water molecules linking both subunits.

This special arrangement between subunits situates their active center cavities confronted in pairs, i.e. the active center of the upper subunit of one dimer is facing the active center of the lower monomer of the dimer placed next (Figure 5(b)). Although the cavities are not completely eclipsed and the space between every pair of confronted cavities is 16 to 20 Å wide, this conformation does not seem very efficient from a catalytic point of view. However, a remarkable feature is the disposition of two residues, both Glu306 of each confronted unit, whose side-chains are clearly protruding towards the entrance of the space enclosed by the two units. These residues might play a role in orientating the substrates in their way to the active centers. It is generally accepted that a greater symmetry usually involves lower catalytic efficiency. Thus, the quaternary structure of BglA might reflect a compromise between the higher degree of aggregation presented by many intracellular proteins and their enzymatic efficiency.

Ligand binding and active center

Glycosyl hydrolases belonging to family 1 are retaining enzymes which are believed to function through a double displacement mechanism in

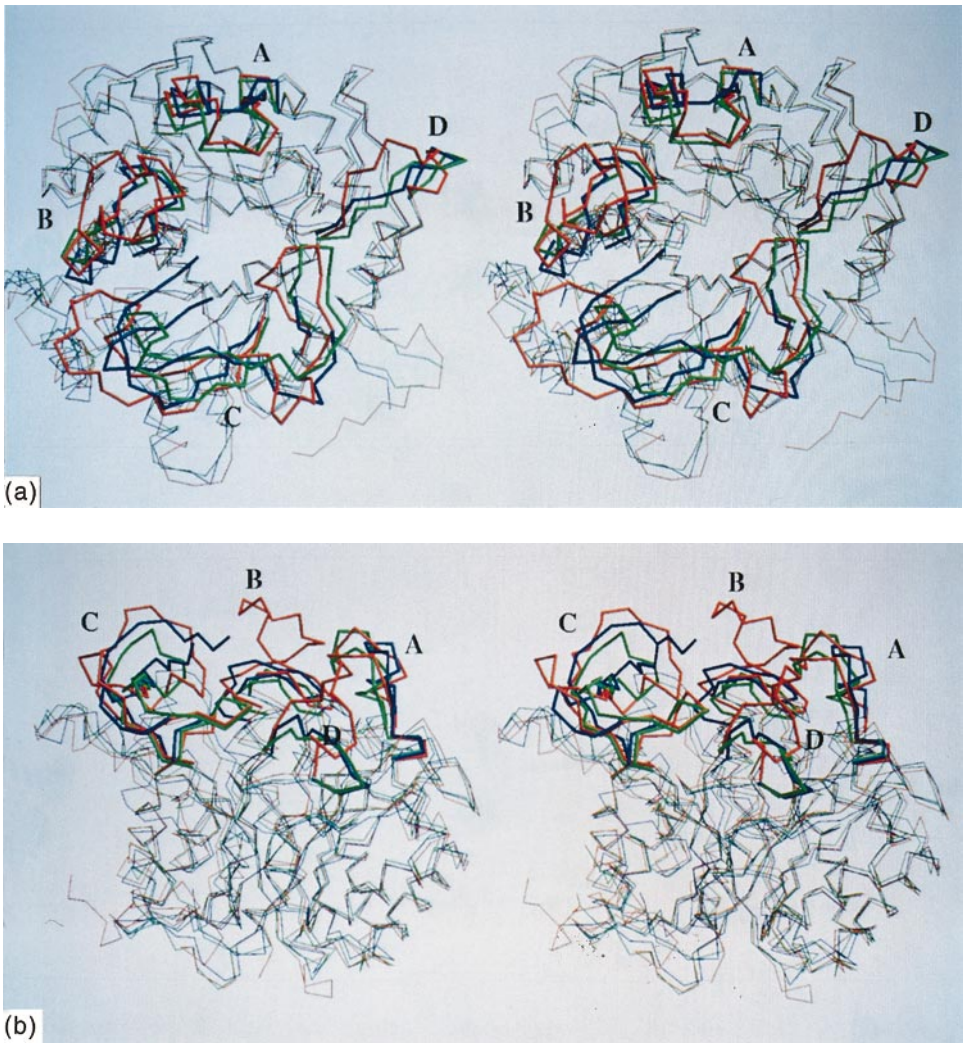

Figure 3. Comparison of monomers in family 1 glycosyl hydrolases with known 3D structure. Stereoview of the C^{α} backbone superposition of BglA (green), 1PBG (blue) and 1CBG (red). Structurally different regions are drawn in thick trace. Loops A, B, C and D are defined in Figure 2. The orientation in (b) is perpendicular to that in (a). Diagram plotted with the program MOLSCRIPT (Kraulis, 1991).

which a glycosyl-enzyme intermediate is generated and hydrolyzed *via* oxocarbonium ion-like transition states (Davies & Henrissat, 1995; McCarter & Withers, 1994; Withers & Aebersold, 1995). In this mechanism, two residues are involved, one acting as nucleophile and the other being an acid/base catalyst. Aldonolactones are well known to be powerful and highly specific inhibitors of the enzymes that hydrolyze the glycosides derived from the related sugar. This effect is attributed to the structural similarity with the glycosyl oxocarbonium ion intermediate, although other electrostatic interactions with the nucleophile could be considered, taking into account the large dipole moment of the lactone (Legler, 1995).

However, aldonolactones are unstable species at pH slightly alkaline, interacting with water and producing complex conformational equilibria which depend on the temperature, pH and age. In particular, D-glucono-1,5-lactone in solution transforms into D-gluconic acid and D-glucono-1,4-lactone (Combes & Birch, 1988). The inhibitor effect is attributed mostly to the 1,5-lactone, inhibition by the 1,4 isomer being explained in terms of the rapid interconversion in aqueous solution. It is generally accepted that the opening of the ring destroys the inhibitory capability of the molecule.

We have co-crystallized BglA with D-glucono-1,5-lactone in the same conditions as the native enzyme, i.e. pH always greater than seven. In these conditions one expects the gluconate species to be predominant, but the inhibitory effect of the 1:5 lactone should be strong enough to displace the equilibrium towards the formation of the 1:5 ligand-enzyme complex. The determination of the crystal structure of this complex unambiguously showed that the ligand was in the open conformation and any attempt to model it in a six or even five-membered ring was completely unsuccessful (Figure 6). Some experiments carried out in the crystallization conditions revealed that the

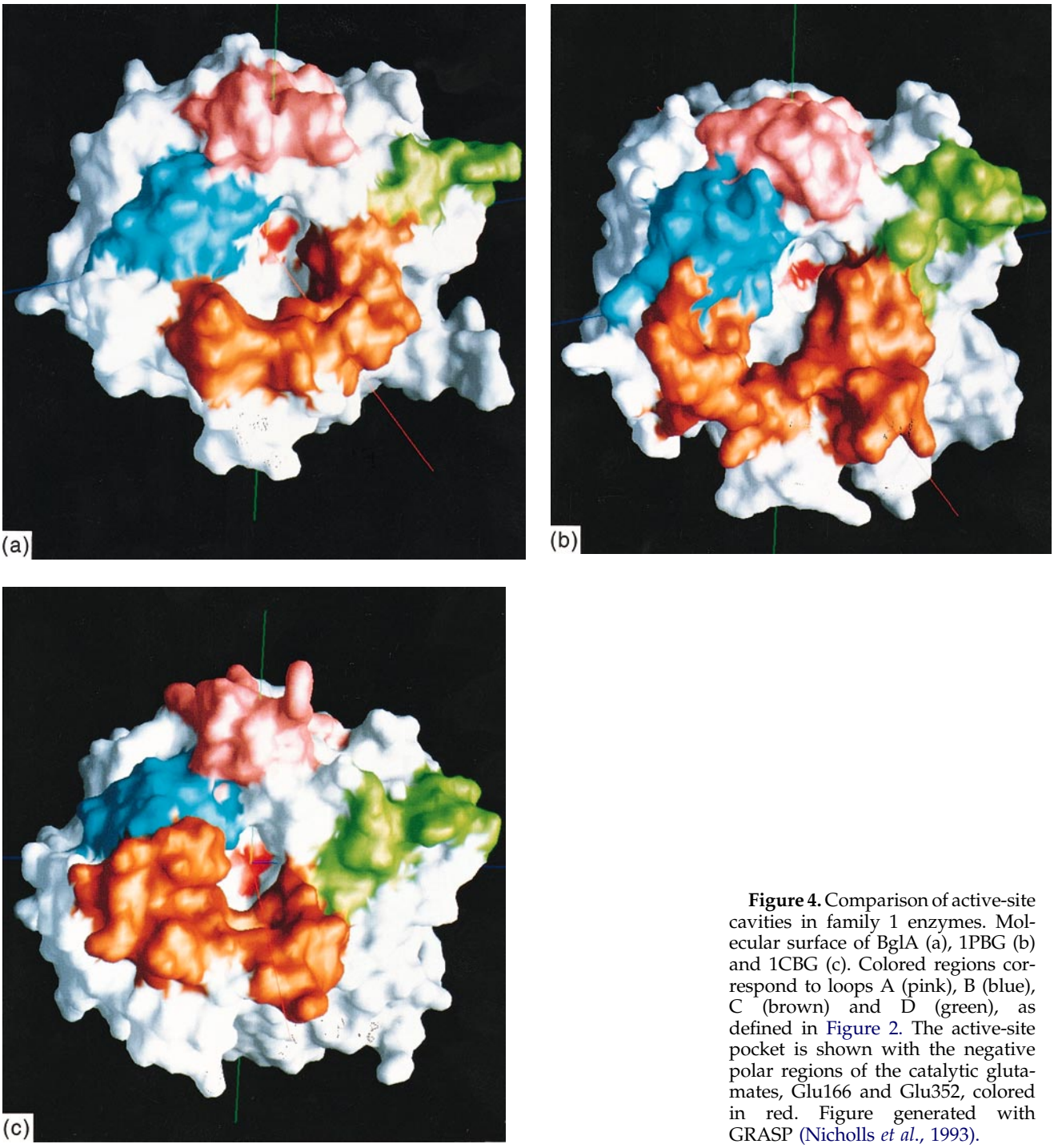

Figure 4. Comparison of active-site cavities in family 1 enzymes. Molecular surface of BglA (a), 1PBG (b) and 1CBG (c). Colored regions correspond to loops A (pink), B (blue), C (brown) and D (green), as defined in Figure 2. The active-site pocket is shown with the negative polar regions of the catalytic glutamates, Glu166 and Glu352, colored in red. Figure generated with GRASP (Nicholls *et al.*, 1993).

enzyme is inhibited only at high concentrations of ligand and that the inhibitor efficiency is comparable to the effect produced by dissolving gluconic acid directly (J. Cañada, personal communication). As shown in Figure 6, the chain adopts a conformation which closely resembles the six-membered ring, and so, mimicking the interaction between the enzyme and the true inhibitor.

The conformation of BglA remains nearly unchanged upon binding of gluconate, as shown by the r.m.s. deviation of 0.2 Å for the main chain between the complexed and the free enzyme. The geometry of the active site is also well conserved, the r.m.s. deviation between residues involved in ligand binding (considering all atoms) being 0.7 Å. The interaction conforms to the typical configuration for sugar-binding sites of proteins, with hydrogen bonds between the sugar hydroxyls and polar side-chains conferring binding affinity and specificity, and aromatic residues oriented against the hydrophobic faces of sugar (Vyas, 1991). In particular Trp398, which is strictly conserved along family 1, is making stacking interaction with the plane defined by the ring-like gluconate chain.

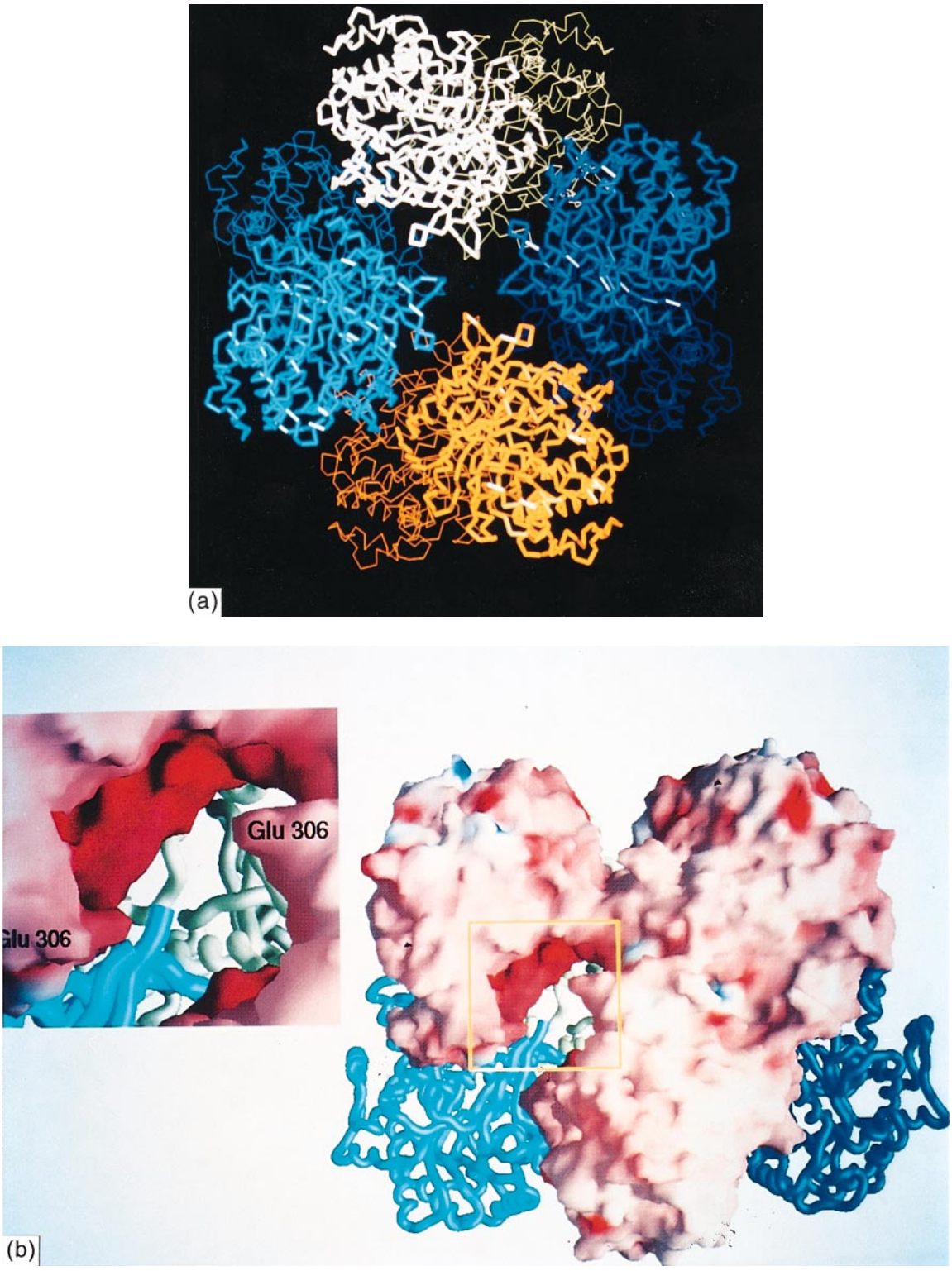

Figure 5. The octameric β-glucosidase A from *Bacillus polymyxa*. (a) C^{α} backbone representation of the octamer as seen along the local 4-fold axis, with crystallographic and local 2-fold axes running horizontally and vertically onto the paper plane. Pairs defining dimers are presented in the same color. (b) A view of the octamer perpendicular to (a) i.e. along the 2-fold axis. Three of the subunits are shown as molecular surfaces with negatively charged zones in red and positively charged in blue. Only the main-chain of the other subunits is shown, keeping the same colors as in (a). Closer interactions are found between pairs related by 2-fold axes, forming a tetramer of dimers. Active site cavities, with marked acidic character, are confronted between different dimers: the active center of the upper subunit of one dimer is faced to the active center of the lower monomer of the dimer placed next. Glu306, whose side-chain is clearly protruding into the hole between cavities, is labeled. A detail of the space between cavities is shown. Figure (a) generated with O (Jones *et al.*, 1991) and (b) with GRASP (Nicholls *et al.*, 1993).

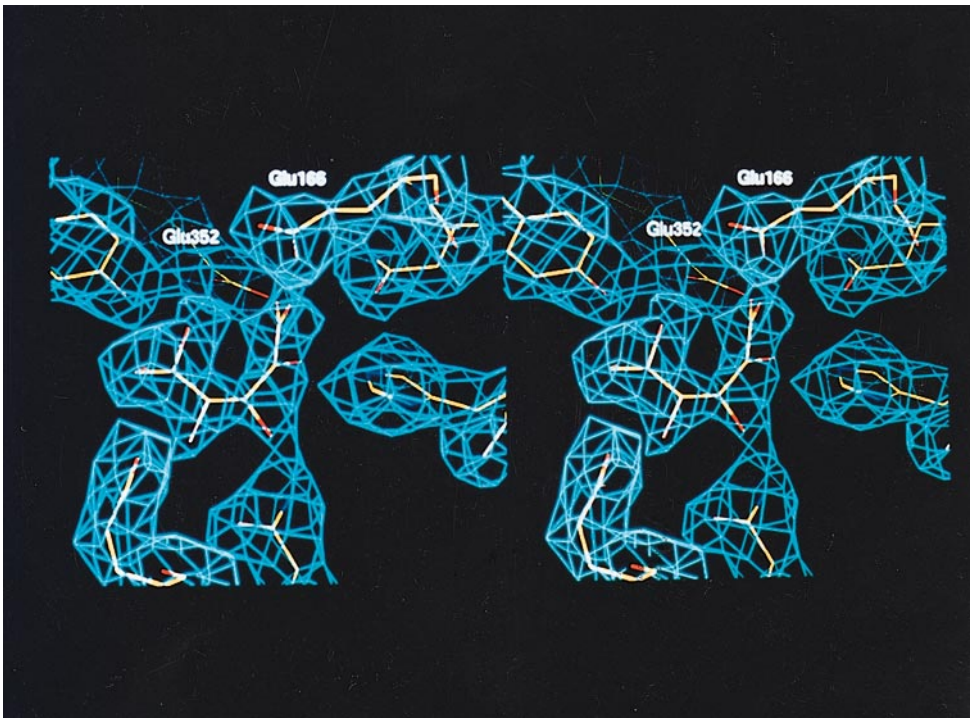

Figure 6. Electron density in the enzyme-gluconate complex. A stereo-view of the active site with the final model for the ligand superimposed into the $(2F_o - F_c)$ contoured at 0.9σ.

Recognition of gluconate by the enzyme is mediated by the residues shown in Figure 7 through a series of hydrogen bonds. Many features of this interaction are shared with the myrosinase glycosyl-enzyme complex (Burmeister *et al.*, 1997), which demonstrates a common pattern in the recognition of *S*- and *O*-glycosides. On the other hand, the location and interactions of the inhibitor in the active site of BglA, confirm the roles proposed for the catalytic residues in family 1 enzymes (Wang *et al.*, 1995; Witt *et al.*, 1993). The acidic residues Glu166 and Glu352 are positioned at a similar distance from the anomeric sugar C1 atom, although the corresponding side-chains approach the substrate in different orientations. The carboxylate group of Glu352 is pointing to the C1 atom in a position consistent with its role in the nucleophilic attack. Glu352 is also making hydrogen bonding with O2 of the gluconate ligand. The hydroxyl substituent at C2 is known to play a crucial role in transition-state stabilization by forming key interactions with the active site. Furthermore, O2 is hydrogen linked to His121 side-chain, which is also conserved along family 1 and was predicted

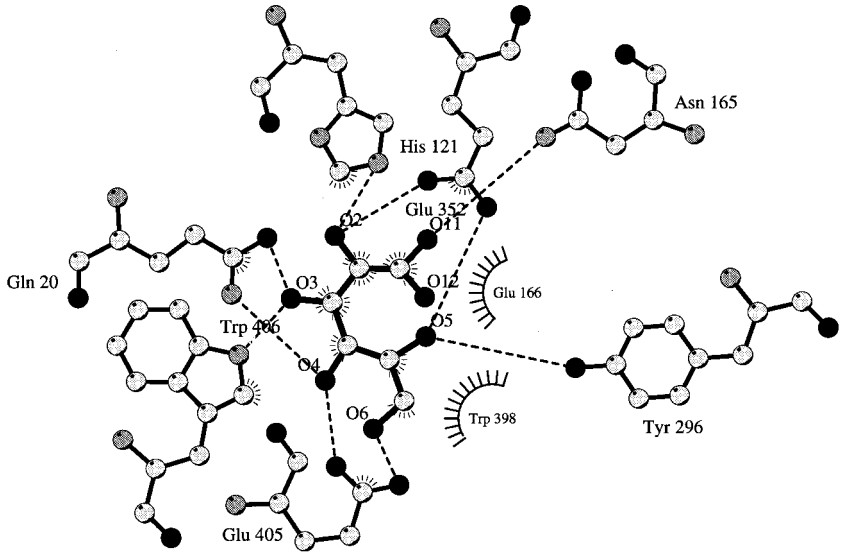

Figure 7. BglA-gluconate interactions. Schematic diagram of the possible enzyme-ligand hydrogen bond network involved in the recognition of the substrate. Plot generated with the program LIGPLOT (Wallace *et al.*, 1995).

to be somehow involved in substrate binding or transition state stabilization (Barrett *et al.*, 1995). Finally, Tyr296 also suspected to be important for catalysis (Wang *et al.*, 1995) and which is hydrogen bonded to Glu352 in the native enzyme, is interacting with O5, and so, it might contribute to stabilization of the oxocarbonium species.

The nitrogen atom of Trp406 is hydrogen-bonded to O3. This residue is conserved along β-glucosidases and phospho-β-galactosidases of family 1, but in most phospho-β-glucosidases is substituted by Ser, or even Cys in ascB of *E. coli*. In the particular case of myrosinases, the corresponding residue is Phe. This possibly means that this interaction is not critical. On the contrary, Gln20 is strictly conserved along the whole family, and makes a bidentate hydrogen bond with O3 and O4 hydroxyl groups. As pointed out by Vyas (1991), bidentate bonds have been reported to be formed by adjacent hydroxyls which are both equatorial or when one is equatorial and the other is axial. Therefore, the enzyme would be able to accommodate substrates derived from the glucose and galactose rings (O4 equatorial and axial, respectively) still conserving the two interactions and with only a slight rearrangement of Gln. This explanation can also be argued to describe the other bidentate bond between Glu405 and oxygen atoms O4 and O6. The glutamate is in the segment predicted as the substrate phosphoryl site (Wiesmann *et al.*, 1995) and is conserved in all the glycosidases of the family 1, but in phospho-glycosidases is replaced by serine. Glutamic acid is able to interact with O4 either in axial or equatorial positions, but the phosphorylated O6 cannot be accommodated. Consequently, the nature of these interactions might conform the structural basis for the high specificity of these enzymes against phosphorylated substrates and the ability of many members of the family to display both, glucosidase and galactosidase activities. This remains to be confirmed by the analysis of the appropiate enzyme-inhibitor complexes and site-directed mutagenesis, which is in progress.

As reported earlier (Kolatkar & Weis, 1996), a limited number of changes are needed in the macrophage asialoglycoprotein receptor to invert the specificity of mannose-binding protein to match the galactose-binding properties of C type lectins. In family 1 glycosidases, subtle changes in the conformational freedom of the involved Glu and Gln residues could possibly make a modified enzyme with a greatly improved Glu/Gal specificity. As stated before, BglA is a convenient reporter enzyme and a very suitable system to study problems related to protein engineering due to its activity being easily assayed by the use of chromogenic substrates. We have undertaken a series of mutagenesis studies in order to establish an increased structural and biochemical data basis which might give a deeper insight into the actual role of the active site residues in the hydrolysis.

Materials and Methods

Enzyme preparation, crystallization and data collection

The isolation and purification of *B. polymyxa* BglA from *E. coli* and its crystallization have been described previously (Sanz-Aparicio *et al.*, 1994). Refinement of the conditions (5 μl of 1.3 M sodium phosphate (pH 8.3) were added to 5 μl of 14 mg/ml protein solution in a hanging drop and equilibrated against 1.3 M phosphate) yielded crystals of larger size, up to 1 mm. For obtaining the enzyme-gluconate complex, BglA was co-crystallized in the presence of the inhibitor which was previously dissolved in 1.3 M phosphate at pH 8.3. Isomorphous crystals were grown as described for uncomplexed BglA, using an enzyme:ligand molar ratio of 1:30. A first data set for the native protein was collected to 3 Å resolution from four crystals on a Mar-Research imaging plate detector at controlled temperature of 5°C, using graphite monochromated CuK_α radiation generated by an Enraf-Nonius rotating anode generator operating at 35 kV/90 mA. These crystals were considered to belong to space group $C222_1$ with lattice parameters $a = 155.6$ Å, $b = 208.3$ Å, $c = 208.7$ Å, and $\alpha = \beta = \gamma = 90°$, four molecules in the asymmetric unit and 70% solvent content (Matthews, 1968). This data set was processed using MOSFLM (Leslie, 1987), ROTAVATA and AGROVATA (Collaborative Computational Project, 1994).

A second data set up to 2.4 Å was collected at the LURE synchrotron station in Orsay (France) using the W32 beamline. A native crystal was soaked in a mother liquor solution containing 30% (v/v) glycerol for a few minutes, introduced into a fiber loop and frozen in a stream of nitrogen of 100 K. The cell dimensions for this crystal at low temperature were determined to be $a = 205.46$ Å, $b = 205.46$ Å, $c = 155.86$ Å, $\alpha = \beta = \gamma = 90°$. The unexpected fact was that the diffraction pattern corresponded in this case to a primitive cell. This could be explained in terms of the reflections with $h + l$ (in the new orientation) odd being systematically much weaker that those with $h + l$ even, an effect which is much more evident for data at resolution lower than 3 Å. Therefore, these reflections were not observed in the data set obtained using the weaker radiation generated by the rotating anode, and thus the pattern could only be indexed as a centered cell. However, we cannot exclude the possibility of cryo-cooling affecting the symmetry. Data for the complexed enzyme were also collected at the LURE station up to 2.3 Å, using a single crystal frozen under the same conditions as the native crystal. Both high-resolution data sets were processed using MOSFLM, ROTAVATA and AGROVATA with *P*422 symmetry.

Structure determination

The BglA structure was solved by molecular replacement using the program AMoRe (Navaza, 1994), on the basis of the 3 Å resolution data set. Coordinates of the β-glucosidase from *T. repens* (PDB code 1CBG) were used as the search model, but truncating to Ala the side-chains of residues not conserved in the sequence alignment, which shows 36% identity. Four solutions corresponding to half an octamer were found, which after rigid body fitting produced an *R* factor of 0.50 and a correlation of 0.27. The octamer presented approximate 422 point group, with the local 4-fold axis parallel to the *a*-axis and the crystallographic 2-fold axis that generates

the octamer passing along *b*-axis, at $\frac{1}{4}$ of *c*-axis. Several simulated annealing refinement cycles performed with X-PLOR (Brünger, 1993) and imposing strict non-crystallographic symmetry between the four molecules of the asymmetric unit led the *R* factor to a value of 0.43 ($R_{free} = 0.50$). At this stage, $(2F_o - F_c)$ and $(F_o - F_c)$ maps were computed and O graphical display package (Jones *et al*., 1991) was used to mutate the Ala residues to the corresponding BglA sequence and to manually build the model into the electron density. The poor quality of these maps, however, did not allow a clear definition of the long loops situated near the C terminus of the chain, in the entrance of the active center cavity, and many doubts about the sequence being correctly placed remained in several points. The model was further refined with the previous X-PLOR protocol to an *R* factor of 0.40 ($R_{free} = 0.47$) after which molecular averaging over the four subunits was performed using the RAVE (Kleywegt & Jones, 1994) set of programs. This technique yielded electron density maps with marked improved quality and allowed unambiguous model building of the protein chain. Positional and grouped *B*-factor refinement maintaining strict non-crystallographic symmetry first, and then relaxing relationships between the four molecules by means of the NCS-restraints algorithm of X-PLOR led to a final *R* factor of 0.27 ($R_{free} = 0.31$), with no water molecules introduced yet. When the high resolution data set was available, the model was re-oriented in the new cell by molecular replacement with AMoRe, getting four solutions in the space group $P4_22_12$. After rigid body fitting the *R* factor was 0.28 and the correlation 0.85 for data to 3.5 Å. The octamer presents again 422 point symmetry and is generated from four molecules by the crystallographic 2-fold axis diagonal to the *ab* plane, which is perpendicular to the non-crystallographic 4-fold axis. The model was subjected to positional and individual restrained *B*-factor refinement, after which visual inspection of electron-density maps was carried out, the better quality of these maps allowing the introduction of minor corrections to the protein model. At the last stage, water molecules were included combined with more rounds of positional and *B*-factor refinement, which led to the final model containing 1559 water molecules and an *R* factor of 0.20 ($R_{free} = 0.27$) for data with $F > 2\sigma$ (*F*) between 8 and 2.4 Å resolution (Table 1). Stereochemistry of the model was checked using PROCHECK (Laskowski *et al*., 1993).

Refinement of BglA-gluconate complex

The initial model for the inhibitor was built from Fourier difference maps using calculated structure factors from the native. Simulated annealing-refined omit maps (Hodel *et al*., 1992) were calculated to confirm the open conformation of the ligand. The omitted region consists of an 8 Å sphere around the substrate region. Crystallographic refinement was carried out as described above for uncomplexed enzyme, assuming full occupancy of bound inhibitor, to which NCS-restraints were not applied. The final refinement parameters for the complex are given in Table 1. Although electron density in molecule A allowed manual building of the inhibitor chain (specially in the computed omit maps), the crystallographic refinement repeatedly yielded poor electron density at this site and so the coordinates of the ligand in molecule A were not included in the final model.

Accession numbers

The coordinates of BglA and its complex with gluconate have been deposited in the Brookhaven Protein Data Bank with entry codes 1BGA and 1BGG, respectively.

Acknowledgments

This work has been funded by Comisión Interministerial de Ciencia y Tecnología, PB93-0120 and ALI-769-94. We thank Dr J. Fontecilla-Camps for helping in the use of cryo-techniques. We also thank Dr Cano for fruitful discussions.

References

Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J. & Hughes, M. A. (1995). The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure*, **3**, 951–960.

Beguin, P. & Aubert, J. P. (1994). The biological degradation of cellulose. *FEMS Microbiol. Rev.* **13**, 25–28.

Brünger, A. T. (1993). *X-PLOR: A system for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S. & Henrissat, B. (1997). The crystal structure of *Sinapsis alba* myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure*, **5**, 663–675.

Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.

Combes, C. L. & Birch, G. G. (1988). Interaction of D-glucono-1,5–lactone with water. *Food Chem.* **27**, 283–298.

Davies, G. & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**, 853–859.

González-Candelas, L., Ramón, D. & Polaina, J. (1990). Sequences and homology analysis of two genes encoding β-glucosidases from *Bacillus polymyxa*. *Gene*, **95**, 31–38.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316.

Henrissat, B. & Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–788.

Henrissat, B. & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695–696.

Henrissat, B. & Romeu, A. (1995). Families, superfamilies and subfamilies of glycosyl hydrolases. *Biochem. J.* **311**, 350–351.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. & Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl Acad. Sci. USA*, **92**, 7090–7094.

Hodel, A., Kim, S. H. & Brunger, A. T. (1992). Model bias in macromolecular crystal structures. *Acta Crystallog. sect. A*, **48**, 851–859.

Jenkins, J., Leggio, L. L., Harris, G. & Pickersgill, R. (1995). β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8−fold α/β architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS Letters,* **362**, 281−285.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A,* **47**, 110−119.

Kempton, J. B. & Withers, S. G. (1992). Mechanism of *Agrobacterium* β-glucosidase: kinetic studies. *Biochemistry,* **31**, 9961−9969.

Kleywegt, G. J. & Jones, T. A. (1994). Halloween...masks and bones. In *From First Map to Final Model* (Bailey, S., Hubbard, R. & Waller, R., eds), pp. 59−66, SERC Daresbury Laboratory, Warrington, UK.

Kolatkar, S. T. & Weis, W. I. (1996). Structural basis of galactose recognition by C type animal lectins. *J. Biol. Chem.* **271**, 6679−6685.

Kraulis, P. J. (1991). MOSLCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946−950.

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.

Legler, G. (1995). Glycoside hydrolases: mechanistic information from studies with reversible and irreversible inhibitors. *Advan. Carbohydr. Chem. Biochem.* **48**, 319−385.

Leslie, A. G. W. (1987). Profile fitting. In *Proceedings of the CCP4 Study Weekend* (Helliwell, J. R., Machin, P. A. & Papiz, M. Z., eds), pp. 39−50, SERC Daresbury Laboratory, Warrington, UK.

López-Camacho, C., Salgado, J., Lequerica, J. L., Madarro, A., Ballestar, E., Franco, L. & Polaina, J. (1996). Amino acid substitution enhancing thermostability of *Bacillus polymyxa* β-glucosidase A. *Biochem. J.* **314**, 833−838.

Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491−497.

McCarter, J. D. & Withers, S. G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**, 885−892.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D Version 2.0: a program for photorealistic molecular graphics. *Acta Crystallog. sect. D,* **50**, 869−873.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157−163.

Nicholls, A., Bharadwaj, R. & Honig, B. (1993). GRASP: graphical representations and analysis of surface properties. *Biophys. J.* **64**, 166.

Painbeni, E., Vallés, S., Polaina, J. & Flors, A. (1992). Purification and characterization of a *Bacillus polymyxa* β-glucosidase expressed in *Escherichia coli*. *J. Bacteriol.* **174**, 3087−3091.

Sanz-Aparicio, J., Romero, A., Martínez-Ripoll, M., Madarro, A., Flors, A. & Polaina, J. (1994). Crystallization and preliminary X-ray diffraction analysis of a type 1 β-glucosidase encoded by the BglA gene of *Bacillus polymyxa*. *J. Mol. Biol.* **240**, 267−270.

Sinnot, M. L. (1990). Catalytic mechanism of enzymic glycosyl transfer. *Chem. Rev.* **90**, 1171−1202.

Vyas, N. K. (1991). Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1**, 732−740.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127−134.

Wang, Q., Trimbur, D., Graham, R., Warren, R. A. & Withers, S. G. (1995). Identification of the acid/base catalyst in *Agrobacterium faecalis* β-glucosidase by kinetic analysis of mutants. *Biochemistry,* **34**, 14554−14562.

Wiesmann, C., Beste, G., Hengstenberg, W. & Schulz, G. E. (1995). The three-dimensional structure of 6-phospho-β-galactosidase from *Lactococcus lactis*. *Structure,* **3**, 961−968.

Withers, S. G. & Aebersold, R. (1995). Approaches to labeling and identification of active site residues in glycosidases. *Protein Sci.* **4**, 361−372.

Withers, S. G., Rupitz, K., Trimbur, D. & Warren, R. A. J. (1992). Mechanistic consequences of mutation of the active site nucleophile Glu358 in *Agrobacterium* β-glucosidase. *Biochemistry,* **31**, 9979−9985.

Witt, E., Frank, R. & Hengstenberg, W. (1993). 6-Phospho-β-galactosidases of Gram-positive and Gram-negative bacteria: comparison of structure and function by kinetic and immunological methods and mutagenesis of the lacG gene of *Staphylococcus aureus*. *Protein Eng.* **6**, 913−920.

Edited by R. Huber

(Received 4 July 1997; received in revised form 2 October 1997; accepted 3 October 1997)