

Structure of Glucoamylase from *Saccharomycopsis fibuligera* at 1.7 Å Resolution

JOZEF ŠEVČÍK,^a ADRIANA SOLOVICOVÁ,^a EVA HOSTINOVÁ,^a JURAJ GAŠPERÍK,^a KEITH S. WILSON^{b,*} AND ZBIGNIEW DAUTER^b
^a*Institute of Molecular Biology, Slovak Academy of Sciences, Bratislava, Slovakia, and* ^b*Department of Chemistry, University of York, York YO1 5DD, England. E-mail: keith@yorkvik.york.ac.uk*

(Received 4 November 1997; accepted 2 February 1998)

Abstract

The yeast *Saccharomycopsis fibuligera* produces a glucoamylase which belongs to sequence family 15 of glycosyl hydrolases. The structure of the non-glycosylated recombinant enzyme has been determined by molecular replacement and refined against 1.7 Å resolution synchrotron data to an *R* factor of 14.6%. This is the first report of the three-dimensional structure of a yeast family 15 glucoamylase. The refinement from the initial molecular-replacement model was not straightforward. It involved the use of an unrestrained automated refinement procedure (uARP) in combination with the maximum-likelihood refinement program *REFMAC*. The enzyme consists of 492 amino-acid residues and has 14 α -helices, 12 of which form an $(\alpha/\alpha)_6$ barrel. It contains a single catalytic domain but no starch-binding domain. The fold of the molecule and the active site are compared to the known structure of the catalytic domain of a fungal family 15 glucoamylase and are shown to be closely similar. The active- and specificity-site residues are especially highly conserved. The model of the acarbose inhibitor from the analysis of the fungal enzyme fits tightly into the present structure. The active-site topology is a pocket and hydrolysis proceeds with inversion of the configuration at the anomeric carbon. The enzyme acts as an exo-glycosyl hydrolase. There is a Tris [2-amino-2-(hydroxymethyl)-1,3-propanediol] molecule acting as an inhibitor in the active-site pocket.

1. Introduction

Glucoamylases (α -1,4-D-glucan glucohydrolases, E.C. 3.2.1.3) are exo-glycosyl hydrolases which catalyse the release of β -glucose from the non-reducing ends of starch and maltooligosaccharides (Pazur & Ando, 1960). They occur in a variety of organisms and show wide variation in catalytic properties such as specific activity, debranching activity (different rates of hydrolysis of 1,4 and 1,6 linkages) and substrate specificity (affinity to high- and low-molecular-weight substrates) (Saha & Zeikus, 1989). Glycosyl hydrolases have been classified into more than 60 families on the basis of sequence similarity (Henrissat, 1991; Henrissat & Bairoch, 1993, 1996; Henrissat & Davies, 1997). To date, all members of

the same glycosyl hydrolase family have been shown to share a common catalytic mechanism. There are excellent overviews of the retaining and inverting mechanisms (McCarter & Withers, 1994) and of the three-dimensional folds (Davies & Henrissat, 1995), and a more extended review of these enzymes has recently been published (Davies, Sinnott *et al.*, 1997). Glucoamylases belong to family 15, and have been well characterized biochemically, mainly through study of fungal and yeast enzymes. The reaction proceeds with inversion of configuration at the anomeric carbon. Glucoamylases are widely used commercially, with a major application in the saccharification of starch to glucose for use in the food and fermentation industry, and hold second position after proteases in the worldwide distribution and sales of industrial enzymes.

To date, for family 15, the three-dimensional structure is known for a proteolytic fragment of a fungal glucoamylase from *Aspergillus awamori* var. X100 truncated to residue 471 from the C-terminus. This fragment contains the complete catalytic domain as well as part of the linker which joins it to the starch-binding domain in the intact enzyme. We refer to this fragment as GluAa throughout the paper. GluAa has been solved in the native form (Aleshin *et al.*, 1992; Aleshin, Hoffman *et al.*, 1994) and in complexes with 1-deoxynojirimycin (Harris *et al.*, 1993), acarbose (Aleshin, Firsov *et al.*, 1994; Aleshin *et al.*, 1996) and D-gluco-dihydroacarbose (Stoffer *et al.*, 1995; Aleshin *et al.*, 1996). These revealed that the glucoamylase catalytic domain has an $(\alpha/\alpha)_6$ barrel fold with the active site in the form of a pocket made up of the connecting segments between the N-termini of the inner helices and the C-termini of the preceding outer helices. The five short stretches which are highly conserved in the 16 known glucoamylase sequences (Coutinho & Reilly, 1994) correspond to the substrate-binding and catalytic regions in the three-dimensional structure of GluAa. Glu179 and Glu400 were identified as the two key catalytic residues, the former acting as a proton donor and the latter as a base (Sierks *et al.*, 1990; Svensson *et al.*, 1990; Harris *et al.*, 1993; Aleshin, Hoffman *et al.*, 1994; Frandsen *et al.*, 1994). The structure of the complex of the starch-binding domain of the *A. niger* enzyme with β -cyclo-dextrin has been determined by NMR (Sorimachi *et al.*, 1997).

The yeast enzymes have no starch-binding domain. Glucoamylase GluSf from the dimorphous yeast *Saccharomycopsis fibuligera* is an extracellular glycoprotein with 492 amino acids and a molecular mass of the fully glycosylated form of 62 kDa. There are no cysteine residues in GluSf, in contrast to GluAa in which there are three disulfide bridges. *N*-glycosidic carbohydrate moieties linked to asparagine make up about

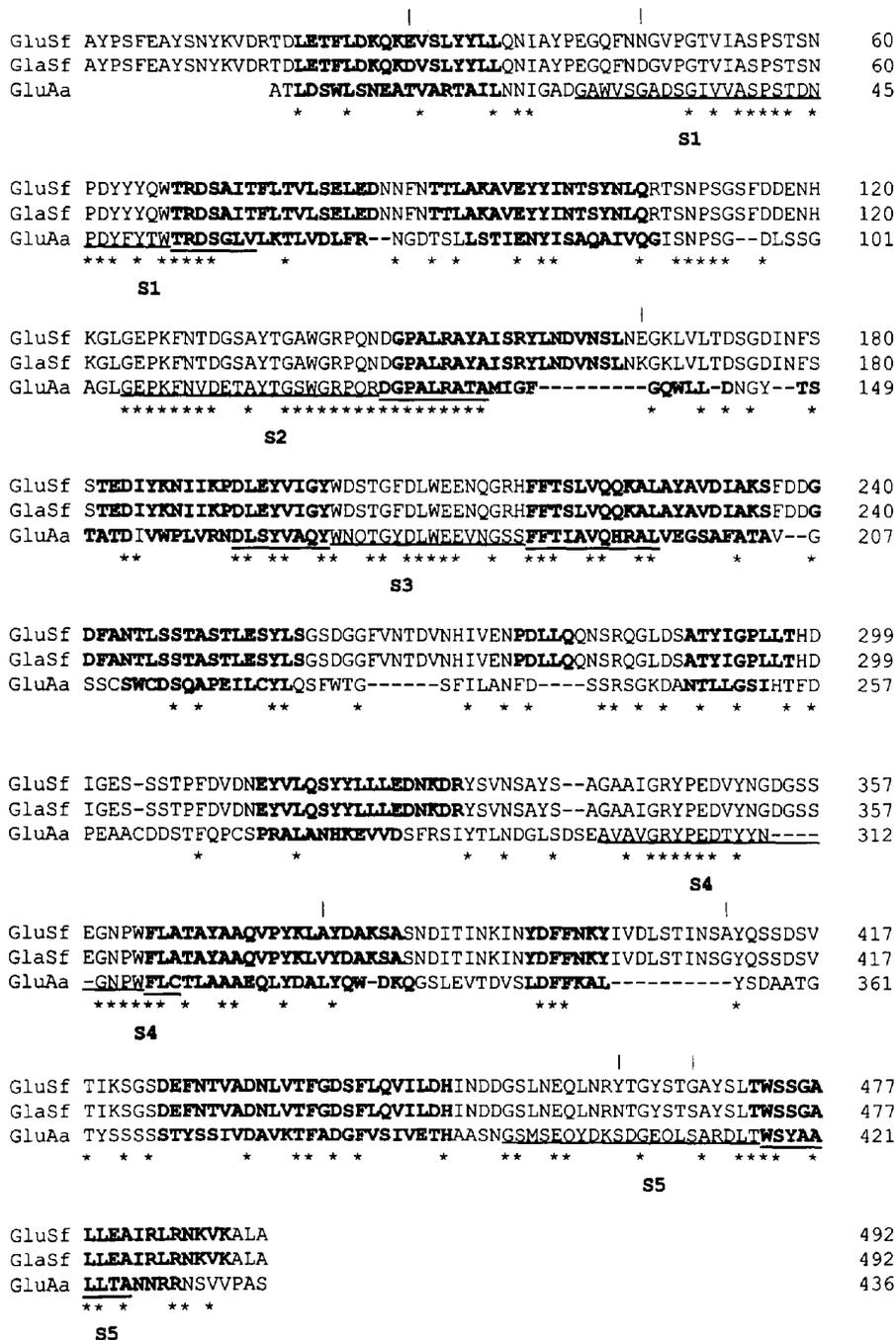


Fig. 1. The primary sequences of GluAa, GlaSf and GluSf. These were aligned using the program *CLUSTALV* (Higgins *et al.*, 1992) and then slightly adjusted manually on the basis of known three-dimensional structures. Residues in the helices are marked in bold. * indicates identities and | indicates differences in the GlaSf and GluSf sequences. There are a number of insertions and deletions, all of which lie within surface loops. The overall identity is 32%. Residues in the five conserved regions in the catalytic domain of glucoamylases defined by Coutinho & Reilly (1994) are underlined.

Table 1. *Data statistics*

Space group	$P2_12_12_1$
Cell parameters (Å)	
<i>a</i>	58.1
<i>b</i>	87.8
<i>c</i>	99.9
Molecules per asymmetric unit	1
Resolution (Å)	29–1.7
Number of unique reflections	56654
Completeness (%)	99.4
R_{merge} (%)	4.2
$I/\sigma(I)$ overall	27.0
$I/\sigma(I)$ (1.71–1.70 Å)	3.5

7 kDa and result in multiple isoforms due to glycosylation heterogeneity (Gašperik *et al.*, 1991). The enzyme is catalytically active up to about 333 K and can regain its activity after total denaturation caused by boiling or high concentration of urea. Two highly homologous genes, *GLU1* (Itoh *et al.*, 1987) and *GLA1* (Hostinová *et al.*, 1991) were isolated from genomic libraries of two variant strains. The updated alignment of their amino-

acid sequences revealed seven amino-acid differences in the mature proteins (Solovicová *et al.*, 1997). These are henceforth referred to as GluSf and GlaSf. The catalytic activity of GluSf is about twice as high as that of GlaSf, the temperature optimum for catalytic activity of both is 325 K and the enzyme is highly active in the pH range 5.2–6.7 (Gašperik & Hostinová, 1993). The sequences of GluSf, GlaSf and GluAa are aligned in Fig. 1. The five conserved regions found in the catalytic domains of 16 glucoamylases (Coutinho & Reilly, 1994) are underlined and residues forming α -helices are shown in bold. The seven amino-acid residues which differ in the sequences of GluSf and GlaSf are marked with a bar.

Some 20 years ago, the glucoamylase from *S. fibuligera* IFO 0111 was employed for commercial production of glucose from starch in Japan (Kato *et al.*, 1976). *S. fibuligera* glucoamylase thus has the potential for commercial exploitation in saccharification of starch during food fermentation, but this will require more extensive studies of the kinetics and physicochemical properties. This structural study forms one step in that analysis.

Crystallization of GluSf from an original producer and recombinant glucoamylase synthesized in *Saccharomyces cerevisiae* was not successful, possibly due to microheterogeneity of the carbohydrate chains. These difficulties were overcome by using enzyme synthesized in *Escherichia coli* (Solovicová *et al.*, 1996, 1997). We report here the 1.7 Å resolution crystal structure of recombinant non-glycosylated glucoamylase corresponding to the *GLU1* gene.

2. Experimental procedures

2.1. Enzyme preparation and crystallization

For crystallization, recombinant non-glycosylated glucoamylase GluSf from the *S. fibuligera* *GLU1* gene overexpressed in *E. coli* was used. Expression and purification of the enzyme from solubilized inclusion bodies preceding *in vitro* renaturation are described in Solovicová *et al.* (1996). Crystallization conditions were as described in Solovicová *et al.* (1997). GluSf crystallizes in space group $P2_12_12_1$, with unit-cell dimensions $a = 58.14$, $b = 87.79$ and $c = 99.95$ Å, from a protein solution of 10 mg ml^{-1} in 50 mM acetate buffer at pH = 4.7–5.5 and 15% (w/v) PEG 8K. Crystals required about two months to grow and the maximum dimensions were approximately $0.3 \times 0.4 \times 0.15$ mm.

2.2. X-ray data collection

X-ray data were collected at room temperature from a single crystal with a MAR Research imaging-plate scanner on the EMBL X31 beamline at DESY, Hamburg. The X-ray wavelength used was 0.93 Å. Two sets of data were measured to 1.7 and 4.0 Å resolution. For both sets a total rotation range of 95° was covered.

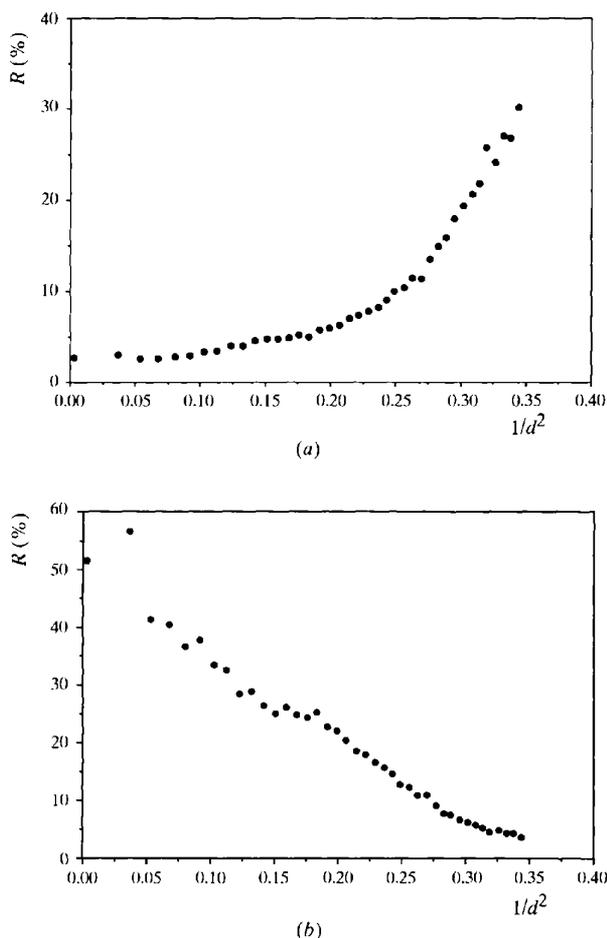


Fig. 2. (a) $R(I)_{\text{merge}}$ as a function of resolution. (b) intensity-to-error ratio, $I/\sigma(I)$.

The data were processed with *DENZO* (Otwinowski & Minor, 1997) and the data statistics are given in Table 1. The merging *R* factor for symmetry-equivalent reflection is shown in Fig. 2(a) and the intensity-to-error ratio (I/σ) in Fig. 2(b).

2.3. Structure determination

The structure was solved by molecular replacement using the program *AMoRe* (Navaza, 1994), with the GluAa catalytic-domain fragment of glucoamylase from *A. awamori* (Aleshin *et al.*, 1992) as a search model [coordinate set 3GLY in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977)]. Both the rotation- and translation-function searches gave a clear and distinct single solution. The rotation function gave a highest (and correct) solution with a correlation coefficient of 18.8 (11.2). The translation function had a correlation

coefficient of 22.8 (8.3) and an *R* factor of 51.8% (55.8%). The figures in brackets refer to the next highest (and incorrect) solutions. Although the solution was clear, the *R* factor was high at over 50%. Rigid-body refinement of the model in the 10–3.5 Å resolution range gave a correlation coefficient of 26.6% and an *R* factor of 51.1%. The next (first wrong) solution was characterized by a correlation of 11.2% and an *R* factor of 55.1%. The properly oriented model was subjected to a series of attempts at standard-refinement protocols using *REFMAC* (Murshudov *et al.*, 1997) combined with *ARP* (Lamzin & Wilson, 1993, 1997) for solvent selection. This approach was unsuccessful. The *R* factor could be reduced to about 35% by releasing the restraints. However, R_{free} remained high at greater than 50% and new interpretable features did not appear in the density maps. This was in spite of an apparently good molecular-replacement solution, with a model whose core was later

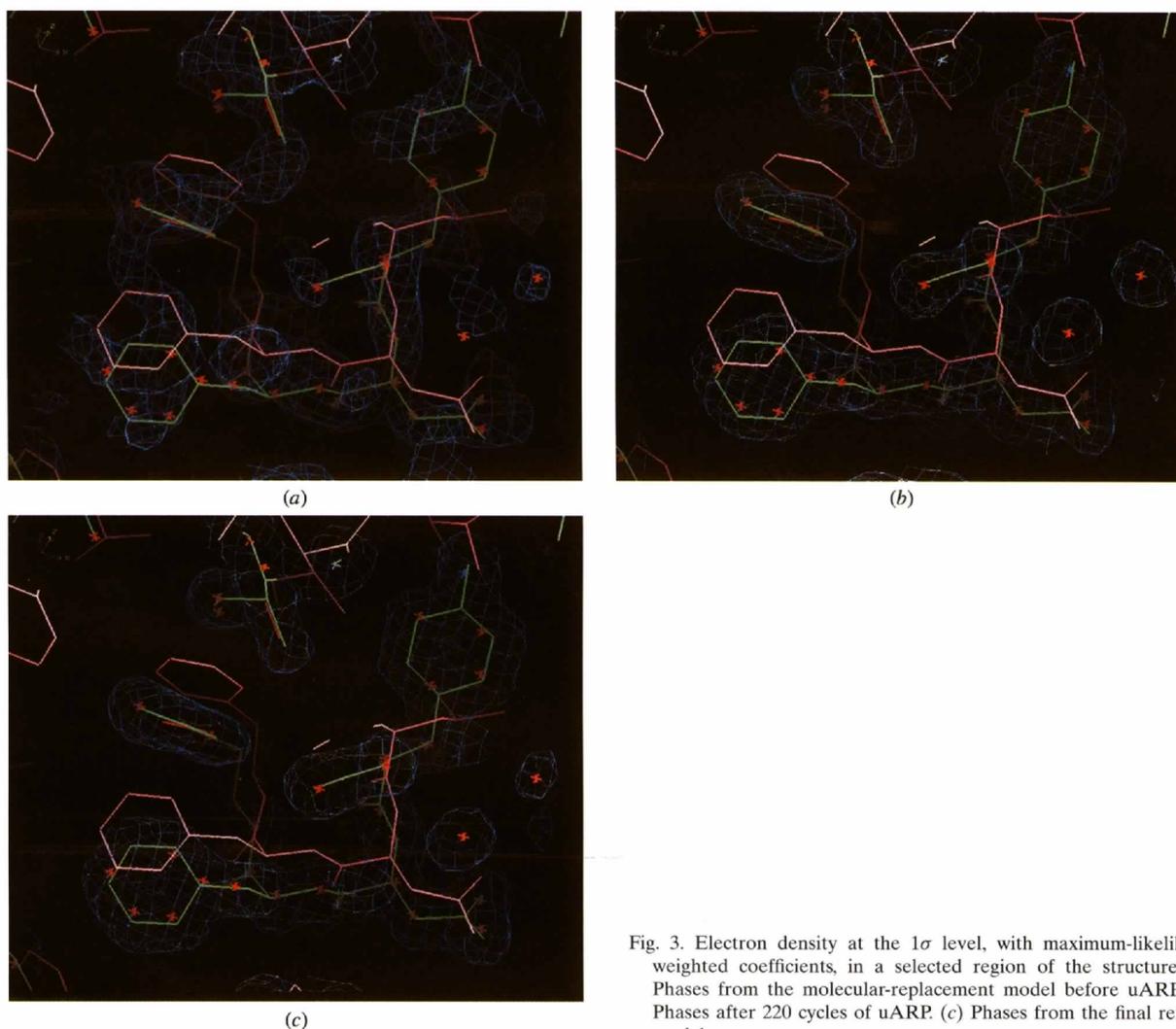


Fig. 3. Electron density at the 1σ level, with maximum-likelihood weighted coefficients, in a selected region of the structure. (a) Phases from the molecular-replacement model before uARP. (b) Phases after 220 cycles of uARP. (c) Phases from the final refined model.

shown to be very similar to the real structure. Presumably an extended series of graphics sessions would have eventually allowed a model to be built, but this would have been a tedious and time-consuming procedure.

Taking into account the high resolution of the data, we embarked upon an alternative procedure involving unrestrained refinement and updating of the atomic model as described by Lamzin & Wilson (1993, 1997).

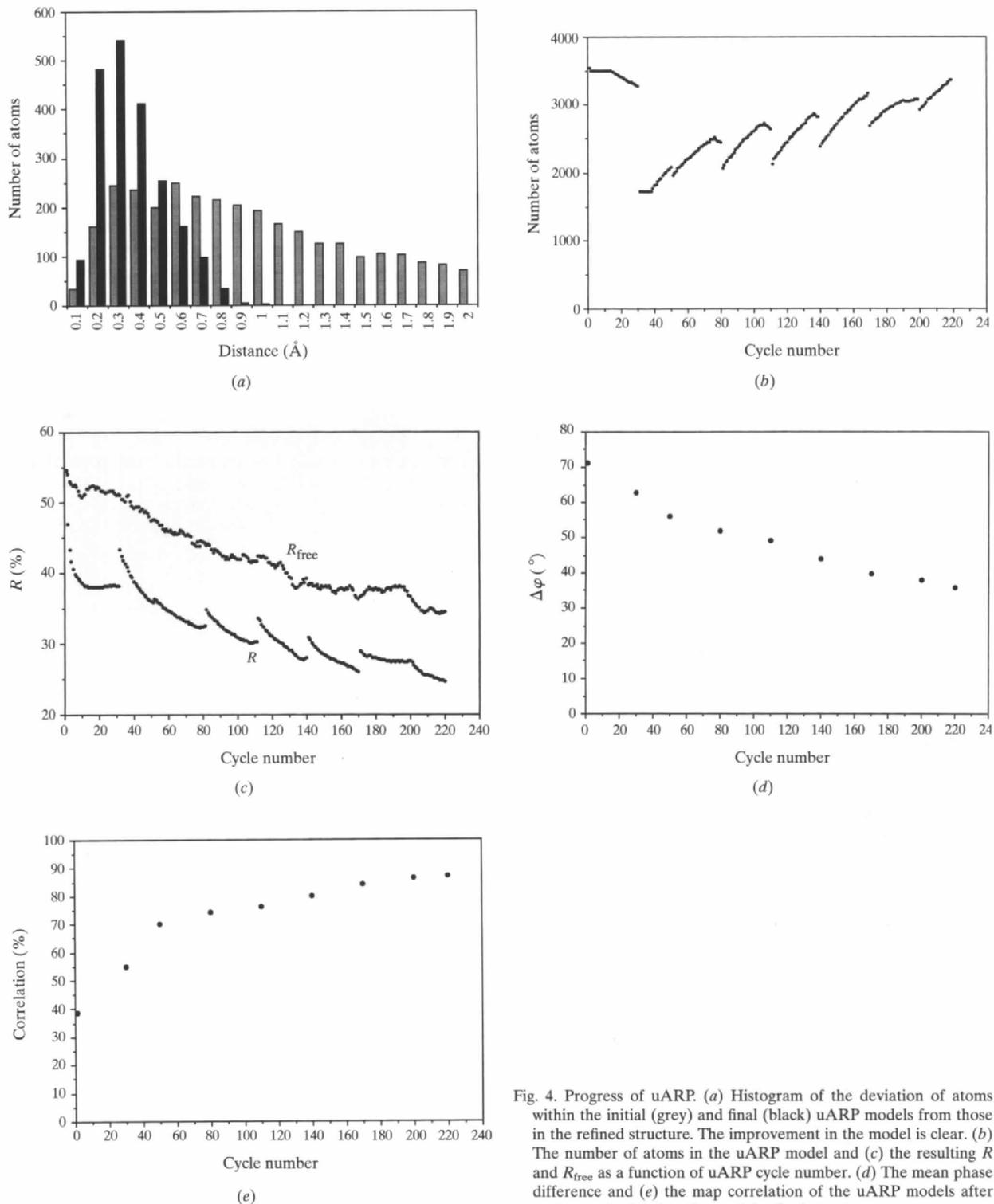


Fig. 4. Progress of uARP. (a) Histogram of the deviation of atoms within the initial (grey) and final (black) uARP models from those in the refined structure. The improvement in the model is clear. (b) The number of atoms in the uARP model and (c) the resulting R and R_{free} as a function of uARP cycle number. (d) The mean phase difference and (e) the map correlation of the uARP models after each round related to those of the final structure.

Table 2. Refinement parameters, restraint weighting scheme and standard deviations after the last refinement cycle

Resolution range (Å)	29–1.7	
<i>R</i> factor (%)	14.6	
<i>R</i> _{free} (%)	17.7	
Number of protein atoms	3898	
Number of solvent atoms	370	
Average <i>B</i> values (Å ²)		
Main chain	16.0	
Side chain	20.7	
Tris	21.4	
Solvent	35.1	
Wilson plot estimate (Å ²)	17.8	
σ_A error estimate (Å)	0.054	
Stereochemistry		Target
Bond distances (Å)	0.021	0.02
Angle distances (Å)	0.037	0.04
Planar 1–4 distances (Å)	0.074	0.05

2.4. Unrestrained ARP

The complete model of GluAa was positioned in the cell according to the molecular-replacement solution. The identities of all 3578 protein atoms were simply redefined as waters. This removed all restraints on stereochemistry and allowed the procedure to explore the conformational space, reducing the effects of prior-model bias and making optimal use of the experimental data. The principles of the minimization were unchanged and the program *REFMAC* was used to optimize the fit between atomic parameters and X-ray observations within the full resolution range, 29–1.7 Å. *REFMAC* was also used to calculate maximum-likelihood weighted terms for map calculations. This provided considerably improved maps, especially in the earlier stages when the *R* factor was high. After each refinement cycle the model was updated according to the following ARP protocol. Up to 30 atoms with the lowest σ -weighted $2mF_o - dF_c$ electron density (if less than 0.5σ) were rejected from the model and up to 30 new atoms were added, based on the difference electron density above 3σ and distances to atoms in the current model in the range 1.0–3.3 Å. After a round consisting of 30 cycles the atoms were ranked according to *B* factor and roughly 25% having the highest *B* values were rejected. The starting value of *R* was 54% and of *R*_{free} 55%. After 220 cycles uARP resulted in an *R* factor of 25% and *R*_{free} of 35%. The whole procedure, somewhat analogous to the Fourier recycling procedure used widely in small-molecule crystallography, required only minimal human intervention and did not involve any intermediate graphics sessions.

The final positions of the water atoms in the uARP model corresponded well to the protein-atom positions, and this procedure provided very good phases, giving high-quality electron density. Examples of the electron density at the beginning of uARP, after 220 cycles with

dummy water positions and after final refinement with the initial and final models are shown in Fig. 3.

The effect of the uARP procedure is shown in Fig. 4(a) which shows histograms of the distances of dummy water molecules from atoms in the final protein molecule in the uARP models at the beginning and after 220 cycles. The course of the uARP is summarized in Figs. 4(b)–4(e): Fig. 4(b) shows the number of atoms in successive uARP models, Fig. 4(c) shows *R* and *R*_{free}, Fig. 4(d) shows the mean phase difference and Fig. 4(e) shows the map correlation after each round in comparison to the final refined values. During the first 30 cycles the number of atoms dropped from about 3600 to 3300 and many acquired high *B* values. Unrestrained ARP in effect retained, or indeed added, only those atoms which corresponded to the GluSf structure. At cycle 30, *R* had only fallen to about 38% and *R*_{free} was still more than 50%. Furthermore, the procedure had essentially converged with little or no fall in *R* and *R*_{free}. This partly explains why the normal restrained-refinement protocol failed. It was therefore decided to delete from the model almost half of the waters, those with highest *B* values, and to continue the procedure. Analogous deletion of up to 500 of the atoms with the highest *B* values (and correspondingly low densities) was applied after every round of ARP, each consisting of 30 cycles (Fig. 4b).

From cycle 30 onwards the number of atoms steadily increased up to roughly the original number, which corresponded to the expected number of atoms in GluSf. The map at around cycle 30 had evidently improved to a sufficient extent to allow uARP to select new atoms correctly positioned in the density and to discriminate against and reject those incorrectly placed. This was reflected in the monotonic fall in *R* and *R*_{free} during each subsequent round (Fig. 4c).

2.5. Model building and final refinement

The effect of unrestrained ARP was dramatic. After uARP the electron density was extremely clear allowing the building of 460 residues out of 492 (93%). The remaining residues were located during the next ten steps, each containing five refinement cycles with more typical restrained ARP. For cross-validation *R*_{free} (Brünger, 1992) was used against the same 2% of reflections as before. Applying an overall anisotropic scaling procedure to the data with *REFMAC* decreased the *R* factor by approximately 0.5%. In the last stages of refinement the H atoms were included at their idealized positions. Refinement with anisotropic scaling, H atoms and all data in the range 29–1.7 Å (56652 X-ray observations) converged with an *R* factor of 14.6% (Table 2).

3. Results and discussion

3.1. Quality of the model

The model of GluSf contains 3871 protein atoms and eight Tris atoms. Double conformations were modelled for the side chains of seven residues: Asp23, Leu82, Val170, Ser251, Leu317, Ser465 and Arg483, leading to a total number of protein atomic sites of 3890. Two proline residues, Pro61 and Pro142, were modelled in the *cis* conformation. There are 370 water molecules all modelled with unit occupancy. The overall coordinate error estimated from the σ_A plot (Read, 1986) is 0.054 Å and the coordinate errors based on R and R_{free} given by *REFMAC* are 0.082 and 0.084 Å, respectively. The average B values for main-chain and side-chain atoms are 16.0 and 20.7 Å², respectively. The average B for Tris is 21.4 Å² and for water molecules 35.1 Å². The overall temperature factor estimated from the Wilson plot (Wilson, 1942) is 17.8 Å² (Table 2). A plot of average thermal parameters as a function of residue number is given in Fig. 5. A small number of residues in loops and

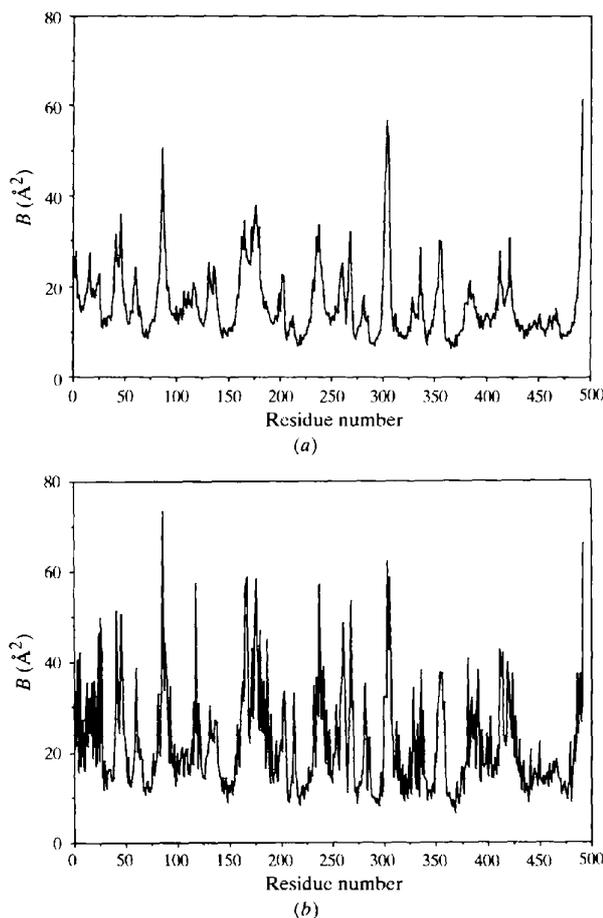


Fig. 5. Average thermal parameters of (a) the main-chain and (b) side-chain atoms as a function of residue number.

at the C-terminus had high B values reflecting substantial flexibility. For some of these, the electron density was very weak and the B values were up to 70 Å². All residues which belong to the active site are in regions with low thermal parameters.

The Ramachandran plot (Ramachandran & Sasisekharan, 1968) calculated with *PROCHECK* (Laskowski *et al.*, 1993) had 91.8% of the residues in most favoured regions and 7.7% in the allowed regions. Only two residues, Ala339 ($\varphi = -117^\circ$, $\psi = -112^\circ$) and Ser357 ($\varphi = -161^\circ$, $\psi = -79^\circ$) lie outside the favoured areas (0.5%), but are nevertheless in the generously allowed regions (Fig. 6). In omit maps, where these residues and their neighbours were removed from the model, the difference electron density confirms that they are correctly placed. Both residues are in surface loops. There is no obvious explanation for the unusual conformational angles of Ala339. However, the CB—OG bond of Ser357 points towards the interior of the molecule and the OG atom forms two hydrogen bonds with Asp352 ND2 (2.97 Å) and the main-chain NH of Glu358 (2.65 Å), which may explain the deviation. Neither residue lies close to the active site.

3.2. The fold of the molecule

The overall fold of GluSf is an $(\alpha\alpha)_6$ barrel as described previously for the proteolytic fragment of the homologous glucoamylase from *A. awamori* (Aleshin *et al.*, 1992). The fold will not therefore be described in

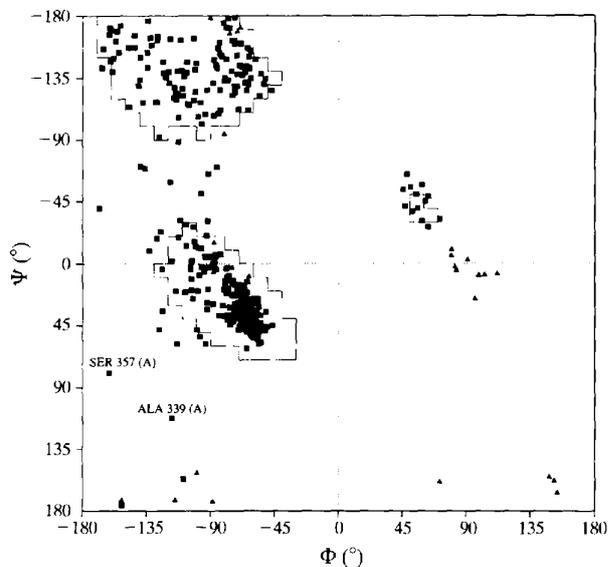


Fig. 6. Ramachandran plot of the refined structure of GluSf produced with the program *PROCHECK* (Laskowski *et al.*, 1993). 99.5% of the residues are in the most-favoured and additionally allowed regions. Two residues are in generously allowed regions. Triangles represent glycines. The contour levels represent the most-favoured and allowed regions as defined in *PROCHECK*.

detail, but we provide a brief overview before comparing it to that of GluAa.

The amino-acid sequence of GluSf contains 492 residues: 230 (46%) are in 14 α -helices, the remainder in loops. There are no β -sheets. 12 of the α -helices make up an $(\alpha/\alpha)_6$ barrel (Fig. 7). The 12 helices form two layers of the barrel. They are organized in six sequential pairs, in which the first member of each pair lies in an outer layer and the second, antiparallel to the first, belongs to an inner layer. Hence, each layer of the barrel is built up of six topologically parallel helices, but the elements of

the two layers are antiparallel to one another. The first and last of these helices, $\alpha 1$ and $\alpha 14$, are antiparallel to one another and complete the barrel. The two extra-short helices, $\alpha 8$ and $\alpha 12$, form protrusions from the connecting loops. Proline residues in the middle of three helices cause kinks, most pronounced in $\alpha 11$ (residues 363–383). The residues composing the helices in GluSf and GluAa are given in Table 3. The core at the centre of the barrel is conical and is filled with hydrophobic side chains, including all but one of the six tryptophans. Both ends of this core are open to solvent. The active site is in

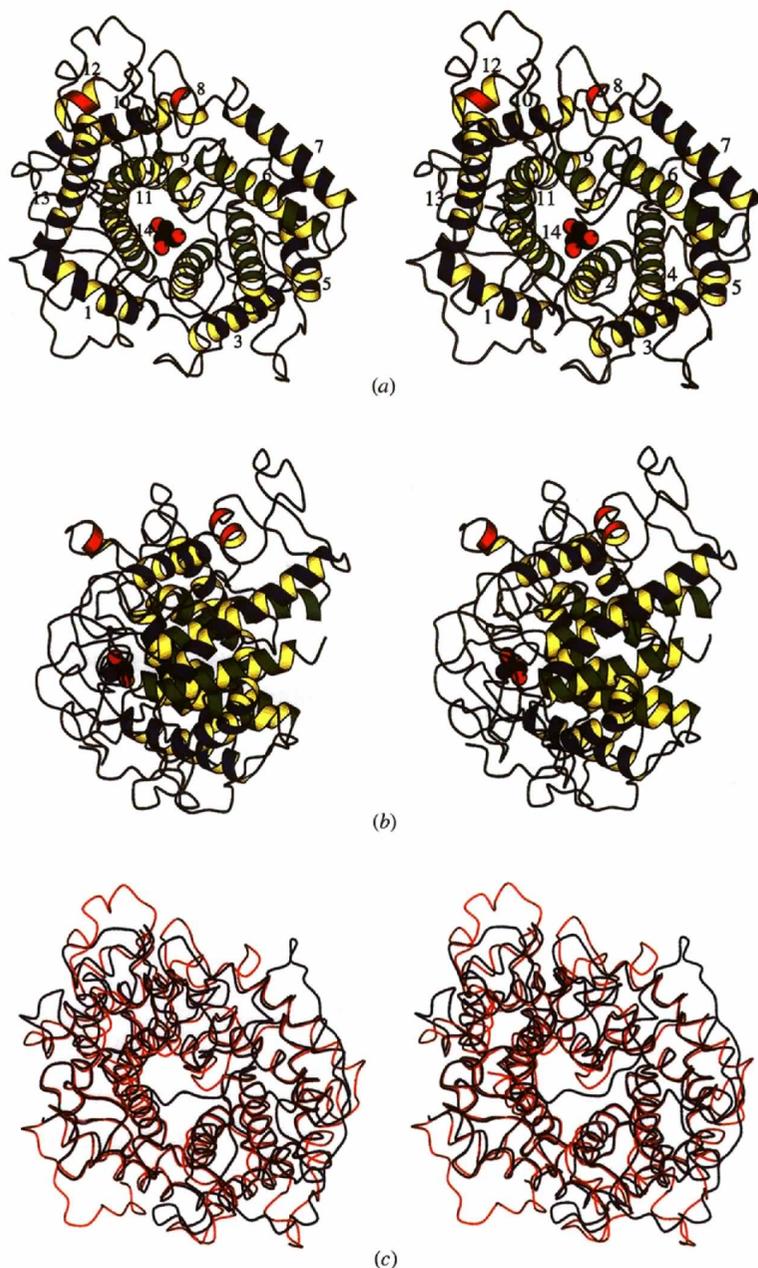


Fig. 7. (a) and (b) Orthogonal stereo ribbon representations of GluSf plotted by *MOLSCRIPT* (Kraulis, 1991). The position of the Tris molecule indicates the active site. (c) Overlap of the GluSf (red) and GluAa (blue) structures.

Table 3. Amino-acid residues forming α -helices in the structures *GluSf* and *GluAa*

	<i>GluSf</i>	<i>GluAa</i>
$\alpha 1$	18–34	3–19
$\alpha 2$	68–84	35–68
$\alpha 3$	89–106	75–90
$\alpha 4$	146–164	126–144
$\alpha 5$	182–200	148–169
$\alpha 6$	217–236	186–205
$\alpha 7$	240–258	211–224
$\alpha 8$	276–280	–
$\alpha 9$	290–298	247–253
$\alpha 10$	314–330	272–283
$\alpha 11$	363–383	318–338
$\alpha 12$	394–400	348–354
$\alpha 13$	424–447	368–391
$\alpha 14$	472–489	417–429

a pocket formed by long loops at the narrower end of the core at the C-terminal end of the external helices or the N-terminal end of the internal ones.

Thus, the fold of this yeast glucoamylase is very similar to that of the catalytic *GluAa* domain of the fungal enzyme. There is no second domain and associated linker in the yeast enzyme. These parts carry the glycosylation sites in the fungal glucoamylase. The reason why the fungal enzyme has, and the yeast enzymes lack, a starch-binding domain is not clear. It may be that the elaborate structures of the natural polysaccharide substrates differ somewhat for the two organisms.

3.3. Comparison to other family 15 enzymes: sequences and three-dimensional fold

The sequences are known for about 19 glycosyl hydrolases from family 15. All can be assumed to share a common catalytic $(\alpha/\alpha)_6$ barrel domain. 16 sequences were previously subdivided into five subfamilies (Coutinho & Reilly, 1994).

(i) Filamentous fungi I (*Aspergillus*). These are around 615 residues long. There are two major domains, the first 440 residues composing the catalytic domain (CD) while roughly 100 C-terminal residues make up a starch-binding domain (SBD) and a linker of about 70 residues joins the two. The linker is extensively glycosylated. The only previously known three-dimensional structure of a glucoamylase is of the catalytic domain of the enzyme from *A. awamori*, *GluAa*, from this subfamily.

(ii) Filamentous fungi II (*Rhizopus*). This subfamily contains only one sequence and is only 30% identical to subfamily (i). There is again an SBD but it is now at the N-terminus of the CD.

(iii) Yeast I (*Saccharomycopsis*). This includes the present *GluSf* and *GlaSf*. These enzymes have 492 residues in a single CD. There is no SBD. They have

about four potential N-glycosylation sites and differ in sequence in only seven positions.

(iv) Yeast II (*Saccharomyces*). These includes two enzymes with a CD but without an SBD. They are most closely related to subfamily (iii), but show some differences towards the C-terminus.

(v) Bacteria (*Clostridium*). Only one sequence is available. It has a long N-terminal extension on the CD via which it is believed to bind the cell wall through a cysteine residue.

The present three-dimensional structure of *GluSf*, from subfamily (iii) is, hence, the first of yeast origin and is representative of a new subclass of these enzymes. The sequences of the catalytic domains of *GluSf* and *GluAa* are only 32% identical. In addition there are a number of insertions and deletions between the two. The similarity of the overall fold of *GluSf* and the CD of *GluAa* is clear from the use of the latter in providing a satisfactory molecular-replacement model. However, the problems in refining the structure from this initial model reflect the substantial differences between the two. The core of the catalytic $(\alpha/\alpha)_6$ barrel is essentially identical in these two enzymes but there are significant differences in the loop regions.

3.4. Mechanism and active site of family 15 hydrolases: current knowledge

Glycoside hydrolysis has been the subject of extensive studies in recent years. The enzymes have been classified into more than 60 families on the basis of sequence homology (Henrissat & Bairoch, 1996). Within each family the enzymes utilize a common mechanism in all examples known to date, although there is some variation in substrate specificity. Differences between families include the use of a retaining or inverting mechanism with regard to the anomeric configuration, exo- or endo-glucoamylase activity, preference for the particular type of glycosidic linkage [e.g. β -(1,4) as opposed to α -(1,4)], selection of the monosaccharide moiety within the substrate and choice of linear or branched saccharides. Inverting glycosyl hydrolases generally utilize two carboxyl groups on the protein directly for catalysis, one acting as an acidic proton donor, the second as a base. The spatial separation of these two functional groups is different for retaining as opposed to inverting enzymes.

The overall properties of family 15 glucoamylases have been clearly established, using a variety of techniques (Natarajan & Sierks, 1996; Sierks & Svensson, 1996). Studies on *GluAa* have been the most extensive and informative and the summary here largely relates to results on that enzyme.

In fungi, yeast and bacteria the glucoamylases are secreted from the cell. They use starch or maltooligosaccharides with α -(1,4) linkages as preferred substrates and can also degrade α -(1,6) linkages but much less

effectively. They are true exoglucanases liberating a single glucose residue from the non-reducing end of the polysaccharide. In addition, under fermentation conditions with reduced water levels, the enzyme is able to condense sugars to form such linkages, which can be a problem in industrial applications. GluAa has at least four binding subsites, which were originally labelled *A* to *D* (Aleshin, Firsov *et al.*, 1994), but are here referred to according to the nomenclature proposed by Davies *et al.* (1997). Hence, they are labelled -1 on the non-reducing side of the scissile bond and $+1$ through $+3$ on the opposite side. The terminal residue is removed by breaking the glycosidic linkage between the glucose units in sites -1 and $+1$. Longer maltooligosaccharides are better substrates for GluAa (Sierks *et al.*, 1989; Sierks & Svensson, 1996) and for the glucoamylase from *Aspergillus niger* (Frandsen *et al.*, 1994). There are at least four sugar-binding subsites, -1 to $+3$, largely identified through the studies of GluAa complexes. The hydrolysis proceeds with inversion of configuration to produce β -D-glucose. The pH optimum for most glucoamylases is in the range 4.0–6.5. Two carboxyl groups act as a proton donor and a base. These are Glu179 and Glu400 in GluAa. The identity of these has been confirmed by a wealth of complementary evidence including the residues conserved in the primary sequences, the three-dimensional structures of GluAa and its complexes and site-directed mutagenesis of these residues in GluAa or homologous enzymes from other species (Coutinho & Reilly, 1994; Svensson *et al.*, 1990; Sierks *et al.*, 1990; Frandsen *et al.*, 1994; Harris *et al.*, 1993). The active site is in the form of a pocket on one end of the $(\alpha/\alpha)_6$ barrel. The non-reducing end of the chain fits tightly into this pocket and this readily explains the exo-glucanase activity of glucoamylases.

3.5. The active site of GluSf

The active site of GluSf is easily identified on the basis of homology to GluAa and by the presence of the Tris molecule. The structures of GluAa (PDB code 3GLY) and GluSf were superimposed using a set of 298 C α atoms from the core of the proteins, mainly within the α -helices. The overlap gives an r.m.s. displacement of 0.76 Å. The C α used in the superposition were selected as those which superimposed within 1.70 Å. As expected, the major differences lie in the loops on the surface. The inner region of the glucoamylase fold is indeed highly conserved. One of the two short α -helices external to the $(\alpha/\alpha)_6$ barrel in GluSf (residues 394–400) is present also in GluAa, but for the other (residues 276–280) the counterpart in GluAa is just a loop. The two catalytic residues are Glu210 (179) and Glu456 (400); the numbers in brackets refer to GluAa.

Essentially all the residues responsible for recognition of the sugar moieties in sites -1 to $+3$ in the structures of

Table 4. Hydrogen bonds (Å) for sugars -1 and $+1$ in the GluAa–acarbose complex compared with the hypothetical GluSf–acarbose complex

The acarbose was kept in the same conformation in the hypothetical model as that observed experimentally in GluAa.

	GluAa		GluSf	
Site -1				
O2	Arg305 NH1	2.99	Arg345 NH1	2.91
O3	Carbonyl 77 O	2.73	Carbonyl 208 O	2.66
O4	Asp55 OD1	2.80	Asp70 OD2	2.87
	Arg54 NH2	2.86	Arg69 NH2	2.99
	Arg54 NE	3.12	Arg69 NE	3.30
O6	Asp55 OD1	2.77	Asp70 OD1	2.86
Site $+1$				
O2	Glu180 OE2	2.68	Glu211 OE2	2.50
O3	Carbonyl 178 O	2.69	Carbonyl 209 O	2.49
	Arg305 NH1	2.74	Arg 345 NH1	2.65
	Glu180 OE2	3.12	Glu211 OE2	3.17
N4	Glu179 OE1	2.66	Glu210 OE1	2.96

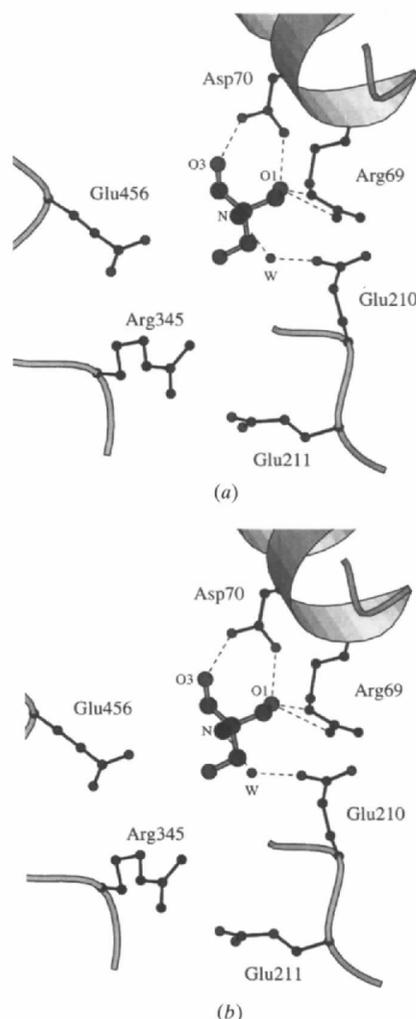


Fig. 8. Hydrogen-bond interactions of Tris [2-amino-2-(hydroxymethyl)-1,3-propanediol] in the -1 site with GluSf. The catalytic residues are Glu210 and Glu456.

the GluAa complexes are conserved in GluSf. For sugar rings -1 and $+1$, the hydrogen-bond contacts are listed in Table 4. The hydrophobic contacts around the same two rings are strictly conserved between GluAa and GluSf, including (with the GluAa sequence numbers in brackets) Tyr63 (48), Trp67 (52), Trp139 (120), Leu208 (177), Trp209 (178), Tyr351 (311), Trp362 (317), Leu471 (415) and Trp473 (417). All the latter have contacts shorter than 4.2 Å between the acarbose in the GluAa complex and the GluSf model (see below). This provides a highly conserved hydrophobic surface on which the sugars rest. Most of these residues are also conserved in the other members of family 15 (Coutinho & Reilly, 1994). Thus, while the experimental data on kinetics and oligosaccharide binding are still to be defined in detail for GluSf, they can be expected to largely parallel those already established for GluAa.

In the void of the barrel there was clear electron density to which a Tris molecule was assigned (Fig. 8). Two of the CH₂OH moieties of the Tris form precise hydrogen-bond interactions with Arg69 and Asp70. The third CH₂OH does not form any hydrogen bonds. The NH₂ group forms hydrogen bonds through a water molecule with Glu210. The Tris molecule in GluSf binds the same amino-acid residues as the -1 residue of acarbose in the complex with GluAa.

Overlap of the structure of the GluAa acarbose complex (PDB code 1GAH) on GluSf based on 289 C α atoms gives r.m.s. deviation 0.75 Å and maximum deviation 1.70 Å. Based on the overlap, acarbose from the complex with GluAa was positioned in the active site of GluSf. It forms the same hydrogen bonds with equivalent residues as in the GluAa complex (Table 4). The amino-acid residues which form hydrogen bonds with acarbose in the complex are shown in Fig. 9. Thus, the yeast and fungal enzymes appear to be extremely

similar in their binding of acarbose. The enzymatic mechanism of the two enzymes can be presumed to be essentially identical.

In the GluAa complexes there is a putative nucleophilic water, Wat500, correctly poised for catalysis, which forms hydrogen bonds to Glu400 OE1 and to the 6OH of the -1 sugar. Overlap with the GluAa complexes reveals no equivalent water in GluSf. This is because of the close proximity (about 2.2 Å) of the C3 atom of the Tris molecule to the position of Wat500 estimated from the GluAa structures.

The presence of a Tris molecule in the active site of GluSf was unexpected. Tris is known to be an inhibitor of glucoamylases (James & Lee, 1996) and of other glycosyl hydrolases such as α -glucosidase (Kimura *et al.*, 1997). Tris can only displace acarbose from GluAa at high concentrations and at a pH above 7.5 (Clarke & Svensson, 1984). Tris buffer was used during purification of GluSf, but was not used in the crystallization medium. The binding constant appears to be sufficient to have carried it through the crystallization. The enzyme prepared in this way is active on its substrate proving that the association constant for the Tris molecule is weaker than that of the substrate.

3.6. Comparison with other $(\alpha/\alpha)_6$ -barrel glycosyl hydrolases

Glycosyl hydrolase families 8 and 9 also have $(\alpha/\alpha)_6$ -barrel folds (Alzari *et al.*, 1996; Juy *et al.*, 1992). The similarity between families 8, 9 and 15 has already been discussed (Alzari *et al.*, 1996). In summary, although the three sets of proteins all have $(\alpha/\alpha)_6$ -barrel topology, the catalytic residues occur on different secondary-structural elements at different positions in the sequence. In

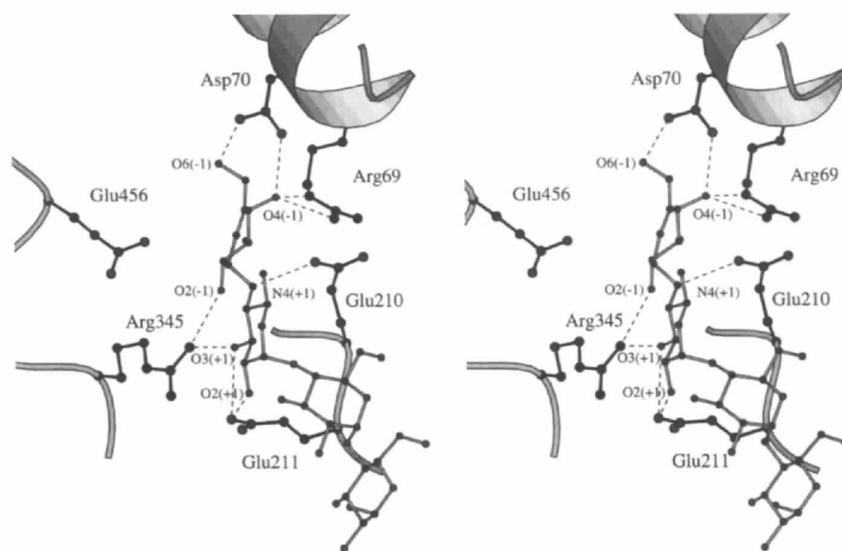


Fig. 9. Hydrogen-bond contacts of acarbose in the hypothetical complex with GluSf, in the same orientation as Fig. 8. The positions of all amino-acid residues are identical with their counterparts in the complex of GluAa with acarbose (1GAH).

addition the specificity sites are quite distinct. Families 8 and 9 have long grooves across the surface (albeit almost perpendicular to one another) consistent with their endocatalytic activity. In contrast the family 15 active site is a pocket, reflecting its exoglucanase activity. The structure of endo/exocellulase E4 from *Thermomonospora fusca* has been recently reported (Sakon *et al.*, 1997). This enzyme is also a member of family 9 and its catalytic domain has the characteristic family fold.

A search of the PDB for similar three-dimensional motifs using the program *SQUID* (Oldfield, 1992), revealed no entries other than those mentioned above with a similar $(\alpha/\alpha)_6$ -barrel fold. However, farnesyl-transferase (Park *et al.*, 1997), which catalyses the carboxy-terminal lipidation of Ras and several other cellular signal-transduction proteins, appears to have a similar topology.

3.7. Comparison with other exoglucanases

The radial nature of the starch granule displays a huge number of non-reducing chain ends. For this reason evolution has provided a number of true exo-enzymes for efficient starch hydrolysis. The glucoamylases described here liberate a monosaccharide product whereas β -amylase liberates the disaccharide maltose in an exo manner. In order for glucoamylase to achieve this with absolute specificity for the terminal sugar, the pyranosyl ring in the buried pocket is entirely surrounded by protein with all the hydroxyl groups of the -1 sugar making direct hydrogen bonds to the protein (Fig. 9).

Comparable active-site pockets have been observed for other exo-glucanases, some of which work on substrates that do not display a large number of free chain-ends in the natural substrate. For example β -galactosidase, β -glucosidase and chitobiase function primarily as exo-disaccharidases. Frequently this is in order to eliminate disaccharide (product) inhibition of endo-enzymes responsible for the primary breakdown of polysaccharides. Others, such as neuraminidase and sialidase, play a key role in viral invasion and bacterial pathogenesis. The key feature of a totally enclosed active-site pocket is conserved in all these exo-enzymes, whether they be inverting or retaining, cleave equatorial or axial glycosidic bonds and even when, as for the most recent example chitobiase (Tews *et al.*, 1996), the nucleophile is donated by the substrate itself.

4. Conclusions

As expected, the yeast glucoamylase has a similar $(\alpha/\alpha)_6$ -barrel structure to the other known family 15 glycosyl hydrolase structures, the catalytic domain of a fungal glucoamylase. Despite the low sequence identity

of about 32% between the sequences of the two enzymes, the residues in the catalytic and specificity sites are almost totally conserved. The acarbose inhibitor from the structure of the fungal-enzyme complex can be easily accommodated in the yeast-enzyme active site.

Future work on this enzyme rests largely on the preparation of mutants and the measurement of detailed kinetic and physicochemical parameters such as stability and activity as functions of pH and temperature, and comparison of these properties with related glucoamylases. This will allow its evaluation in potential industrial applications. In addition the crystallographic studies will be extended to higher, ideally atomic, resolution for the native enzyme and a series of complexes. Diffraction from the present crystals clearly shows that they are of high quality and will provide data well beyond 1.7 Å. The comparison with other family 15 enzymes, such as GlaSf, whose structure determination is under way, is a key aspect of this work.

In terms of crystallographic methodology, the use of unrestrained ARP to refine a molecular-replacement model which was sufficiently different from the real structure to prevent a more simplistic approach, should be useful for other proteins. The fungal enzyme is only 32% identical in sequence to the yeast, and although it provided a ready solution to the rotation and translation functions, direct use of restrained maximum-likelihood refinement to proceed to the correct minimum was unsuccessful. In contrast, removal of the stereochemical restraints using ARP with the *REFMAC* maximum-likelihood program, allowed the initial atoms to migrate, through rejection and acceptance of new atoms, to the positions of real atoms in the yeast structure. The resulting unrestrained ARP model gave phases much closer to those of the refined structure, and an electron-density map into which the structure could be easily built.†

The authors thank Dr Tetsuya Itoh for providing us with the *GLU1* gene, the Slovak Academy of Sciences for grants 2/1070 and 2/1076 and EMBL Hamburg for synchrotron beam time. This work was supported by the Howard Hughes Medical Institute grant 75195-547601 and the Slovak Academy of Sciences grant 2/1070. We acknowledge the support of EC Copernicus contract No. CIPA-CT94-0198 for JS to visit Hamburg and York, and of HHMI grant 75195-547601 for ZD and KSW to visit Bratislava. We thank Dr Gideon Davies for critical reading of the manuscript.

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1AYX).

References

- Aleshin, A. E., Firsov, L. M. & Honzatko, R. B. (1994). *J. Biol. Chem.* **269**, 15631–15639.
- Aleshin, A. E., Golubev, A., Firsov, L. M. & Honzatko, R. B. (1992). *J. Biol. Chem.* **267**, 19291–19298.
- Aleshin, A. E., Hoffman, C., Firsov, L. M. & Honzatko, R. B. (1994). *J. Mol. Biol.* **238**, 575–591.
- Aleshin, A. E., Stoffer, B., Firsov, L. M., Svensson, B. & Honzatko, R. B. (1996). *Biochemistry*, **35**, 8319–8328.
- Alzari, P. M., Souchon, H. & Dominguez, R. (1996). *Structure*, **4**, 265–275.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Simanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 2055–2064.
- Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
- Clarke, A. J. & Svensson, B. (1984). *Carlsberg Res. Commun.* **49**, 559–566.
- Coutinho, P. M. & Reilly, P. J. (1994). *Protein. Eng.* **7**, 393–400.
- Davies, G. & Henrissat, B. (1995). *Structure*, **3**, 853–859.
- Davies, G., Sinnott, M. L. & Withers, S. G. (1997). *Comprehensive Biological Catalysis*, Vol. 1, edited by M. L. Sinnott, pp. 119–209. New York: Academic Press.
- Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem J.* **321**, 557–559.
- Frandsen, T. P., Dupont, C., Lehmbeck, J., Stoffer, B., Sierks, M. R., Honzatko, R. B. & Svensson, B. (1994). *Biochemistry*, **33**, 13808–13816.
- Gašperík, J. & Hostinová, E. (1993). *Curr. Microbiol.* **27**, 11–14.
- Gašperík, J., Kováč, L. & Mináriková, O. (1991). *Int. J. Biochem.* **23**, 21–25.
- Harris, E. M. S., Aleshin, A. E., Firsov, L. M. & Honzatko, R. B. (1993). *Biochemistry*, **32**, 1618–1626.
- Henrissat, B. (1991). *Biochem. J.* **280**, 309–316.
- Henrissat, B. & Bairoch, A. (1993). *Biochem. J.* **293**, 781–788.
- Henrissat, B. & Bairoch, A. (1996). *Biochem. J.* **316**, 695–696.
- Henrissat, B. & Davies, G. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.
- Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992). *Comput. Appl. Biosci.* **8**, 189–191.
- Hostinová, E., Balanová, J. & Gašperík, J. (1991). *FEMS Microbiol. Lett.* **83**, 103–108.
- Itoh, T., Ohtsuki, I., Yamashita, I. & Fukui, S. (1987). *J. Bacteriol.* **169**, 4171–4176.
- James, J. A. & Lec, B. H. (1996). *Biotech. Lett.* **18**, 1401–1406.
- Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeysens, M., Beguin, P. & Aubert, J.-P. (1992). *Nature (London)*, **357**, 89–91.
- Kato, K., Kuswanto, K., Banno, I. & Harada, T. (1976). *J. Ferment. Technol.* **54**, 831–837.
- Kimura, A., Takata, M., Fukushi, Y., Mori, H., Matsui, H. & Chiba, S. (1997). *Biosci. Biotechnol. Biochem.* **61**, 1091–1098.
- Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
- Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* **D49**, 129–147.
- Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- McCarter, J. D. & Withers, S. G. (1994). *Curr. Opin. Struct. Biol.* **4**, 885–892.
- Murshudov, G. N., Vagin, A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Natarajan, S. K. & Sierks, M. R. (1996). *Biochemistry*, **35**, 15269–15279.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Oldfield, T. J. (1992). *J. Mol. Graph.* **10**, 247–252.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Park, H.-W., Boduluri, S. R., Moomaw, J. F., Casey, P. J. & Beese, L. S. (1997). *Science*, **275**, 1800–1804.
- Pazur, J. H. & Ando, T. (1960). *J. Biol. Chem.* **235**, 297–302.
- Ramachandran, S. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–437.
- Read, R. J. (1986). *Acta Cryst.* **A42**, 140–149.
- Saha, B. C. & Zeikus, J. G. (1989). *Starch*, **41**, 57–64.
- Sakon, J., Irwin, D., Wilson, D. B. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 810–818.
- Sierks, M. R., Ford, C., Reilly, P. J. & Svensson, B. (1989). *Protein Eng.* **2**, 621–625.
- Sierks, M. R., Ford, C., Reilly, P. J. & Svensson, B. (1990). *Protein Eng.* **3**, 193–198.
- Sierks, M. R. & Svensson, B. (1996). *Biochemistry* **35**, 1865–1871.
- Solovicová, A., Gašperík, J. & Hostinová, E. (1996). *Biochem. Biophys. Res. Commun.* **224**, 790–795.
- Solovicová, A., Gašperík, J., Ševčík, J. & Hostinová, E. (1997). *Acta Cryst.* **D53**, 782–783.
- Sorimachi, K., Le Gal-Coëffet, M.-F., Williamson, G., Archer, D. B. & Williamson, M. P. (1997). *Structure*, **5**, 647–661.
- Stoffer, B., Aleshin, A. E., Firsov, L. M., Svensson, B. & Honzatko, R. B. (1995). *FEBS Lett.* **358**, 57–61.
- Svensson, B., Clarke, A. J., Svendsen, I. & Moller, H. (1990). *Eur. J. Biochem.* **188**, 29–38.
- Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S. & Vorgias, C. E. (1996). *Nature Struct. Biol.* **3**, 638–648.
- Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.