Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the novel α-amylase AmyC from *Thermotoga maritima*

Achim Dickmanns,[a] Meike Ballschmiter,[b] Wolfgang Liebl[b] and Ralf Ficner[a]*

[a]Abteilung Molekulare Strukturbiologie, Institut für Mikrobiologie und Genetik, Georg-August-Universität, Justus-von-Liebig Weg 11, 37077 Göttingen, Germany, and [b]Abteilung Angewandte Mikrobiologie, Institut für Mikrobiologie und Genetik, Georg-August-Universität, Grisebachstrasse 8, 37077 Göttingen, Germany

Correspondence e-mail: rficner@gwdg.de

Received 26 October 2005
Accepted 9 December 2005

PDB Reference: AmyC, 2b5d, r2b5dsf.

α-Amylases are essential enzymes in α-glucan metabolism and catalyse the hydrolysis of long sugar polymers such as amylose and starch. The crystal structure of a previously unidentified amylase (AmyC) from the hyperthermophilic organism *Thermotoga maritima* was determined at 2.2 Å resolution by means of MAD. AmyC lacks sequence similarity to canonical α-amylases, which belong to glycosyl hydrolase families 13, 70 and 77, but exhibits significant similarity to a group of as yet uncharacterized proteins in COG1543 and is related to glycerol hydrolase family 57 (GH-57). AmyC reveals features that are characteristic of α-amylases, such as a distorted TIM-barrel structure formed by seven β-strands and α-helices (domain A), and two additional but less well conserved domains. The latter are domain B, which contains three helices inserted in the TIM-barrel after β-sheet 2, and domain C, a five-helix region at the C-terminus. Interestingly, despite moderate sequence homology, structure comparison revealed significant similarities to a member of GH-57 with known three-dimensional structure, *Thermococcus litoralis* 4-glucanotransferase, and an even higher similarity to a structure of an enzyme of unknown function from *Thermus thermophilus*.

1. Introduction

Starch, glycogen and related carbohydrates play an essential role in the metabolism of animals, plants and microorganisms. Numerous different sugar-metabolizing enzymes have evolved that catalyze various reactions in the biosynthesis and degradation of carbohydrates. Within the carbohydrate-degradation pathway, α-1,4-glucan-4-glucanohydrolases (α-amylases, EC 3.2.1.1) catalyse the hydrolysis of α-D-(1,4)-glycosidic linkages in linear and branched polysaccharides. Catalysis by α-amylases involves the protonation of the acetal linkage between the *n*-sugar residue and the rest of the polysaccharide chain, resulting in a covalent glycosyl intermediate whose formation and breakdown is accomplished *via* an oxocarbenium-ion-like transition state (Davies *et al.*, 1997; Zechel *et al.*, 1998; Zechel & Withers, 2000). Subsequently, a glucose oligosaccharide is released that can be converted into free glucose and, in downstream processes, to energy. Moreover, α-amylases have become one of the most valuable enzymes in biotechnology, with particular application in the conversion of starch, in baking and in the pharmaceutical industry (Nielsen & Borchert, 2000).

The three-dimensional structures of approximately 130 α-amylases have been determined so far and the amino-acid residues involved in catalysis have been studied extensively (for reviews, see Horvathova *et al.*, 2001; Svensson, 1994).

In general, α-amylases consist of three domains called A, B and C. Domain A is a TIM barrel which makes up the central

part of the molecule. Domain *B* is an irregular β-strand structure which is a protrusion between the third strand and the third helix in the TIM barrel. Domain *C* contains a Greek-key motif and is positioned on the opposite side of the TIM barrel with respect to domain B.

The genome of the strictly anaerobic *Thermotoga maritima* was one of the first organisms to be fully sequenced (Nelson *et al.*, 1999) and has also been chosen for a structural genomics project (JCSG; Joint Centre for Structural Genomics). *T. maritima* is a hyperthermophilic organism which grows anaerobically in hot springs with a growth temperature range of 328–363 K (optimum at 353 K) and a wide pH range (5.5–9.0) and feeds on a variety of saccharides such as glucose, saccharose, starch or xylan. So far, two α-amylases have been described in *T. maritima* (Liebl *et al.*, 1997; Lim *et al.*, 2003). Interestingly, *T. maritima* encodes at least one additional α-amylase, designated AmyC, which was identified by screening a *T. maritima* library for recombinant *Escherichia coli* clones with thermostable amylolytic activity. The corresponding gene (ORF tm1438) is found in an unusual genetic context suggesting a cytoplasmic function. In the published genome sequence of *T. maritima*, the AmyC-encoding ORF is interrupted by an 'authentic frameshift' (Nelson *et al.*, 1999), resulting in a stop codon at bp 526, a subsequent second start codon at bp 555 and a corresponding stop at bp 1585, leading to two genes. Resequencing showed that the stop codon was absent in the sequence and that the frameshift is not present (Ballschmiter *et al.*, 2006). Thus, the now 1587 bp ORF tm1438 encodes full-length AmyC of 528 amino acids in length, resulting in a calculated molecular weight of 62.8 kDa.

Comparisons of the full-length sequence of AmyC with the database revealed a moderately significant similarity of the N-terminal three-fifths of AmyC to members of glycoside hydrolase family 57 (GH-57; Pfam03065), which is a diverse family composed of at least seven subgroups (Zona *et al.*, 2004) and so far 79 members (according to the CAZy database; only 23 members according to Pfam03065). This family includes α-amylase (EC 3.2.1.1), amylopullulanase and 4-glucanotransferase (EC 2.4.1.–) enzymes. For example, the α-glucanotransferase from *Thermococcus litoralis* is well characterized and its structure has been solved (Imamura *et al.*, 1999, 2001, 2003; Jeon *et al.*, 1997). Most of the family 57 enzymes mediate reactions similar to those of α-amylase family members (GH-13 and the related members GH-70 and GH-77). However, no significant similarity has been detected at the amino-acid sequence level between GH-57 and the α-amylase family enzymes. More interestingly, AmyC is closely related to the COG1543-domain proteins (clusters of orthologous groups of proteins), a group of proteins with a so far unknown biochemical function which correspond to almost the full-length sequence of AmyC (Ballschmiter *et al.*, 2006). Less significantly, the COG1449-domain proteins, comprising mainly α-amylases and α-mannosidases, reveal a homology to AmyC but lack the N-terminal region. The COGs have been delineated by comparing protein sequences encoded in 43 complete genomes, including *T. maritima*, representing major phylogenetic lineages (Tatusov *et al.*, 1997, 2003).

The fact that AmyC reveals sequence similarity to the COG1543 proteins and, although to a lesser extent, to GH-57 as well as the lack of similarity between the sequences of known canonical α-amylases and the novel *T. maritima* α-amylase AmyC prompted us to determine the crystal structure of AmyC. Similar to other known amylases, AmyC is composed of three structural domains: domain A, a distorted TIM-barrel structure with a characteristic sevenfold (β-strand/α-helix) motif, domain B, a three-helix insertion that interrupts the TIM barrels after β-strand 2 and before α-helix 2, and domain C, a five-helix region at the C-terminus. Structure comparison revealed differences from the known crystal structure of a member of GH-57, *T. litoralis* 4-glucanotransferase, except for the highly similar TIM barrel in domain *A*, which shows a similar structural arrangement of the catalytically active residues. Interestingly, an even higher similarity was found to a structure of a protein of unknown function from *Thermus thermophilus* (PDB code 1ufa) crystallized by a structural genomics project.

2. Materials and methods

2.1. Cloning

The coding region of the intracellular *T. maritima* MSB8 amylase AmyC was cloned into the overexpression vector pET-24c (Novagen, Schwalbach, Germany) *via* *Nde*I and *Bam*H1 restriction sites introduced at the ends of the gene by PCR. The correct cloning of the ORF was confirmed by both sequencing and restriction-enzyme analysis. The *E. coli* strains BL21(DE3) and the methionine-auxotrophic strain B834(DE3) (Novagen) were transformed with the resulting plasmid pET24-AmyC through electroporation.

2.2. Expression

2.2.1. Wild-type AmyC. An overnight culture of transformed BL21(DE3) cells was grown in LB medium and used to inoculate a 1 l culture of fresh medium in a 1:20 ratio and incubated at 303 K. At an OD_{600} of 0.8, the cells were induced with 0.1 m*M* IPTG. After 6 h expression, the cells were harvested by centrifugation at 6000*g* and 277 K for 20 min, washed in 20 m*M* Tris pH 8.0 and finally resuspended in 6 ml of the same buffer.

2.2.2. Selenomethionine labelling of AmyC. A 1 l culture of *E. coli* B834(DE3) was grown in M9 mineral medium (Reuter *et al.*, 1999) containing 25 mg kanamycin and 0.4%(*w*/*v*) glucose as the carbon source at 303 K in a 5 l baffle flask. The medium was further supplemented with 50 mg L-methionine, 2 mg biotin and 2 mg thiamine. On reaching an OD_{600} of 0.8, after 3.5 h growth following inoculation [2%(*v*/*v*)] from an overnight pre-culture in M9 medium, the cells were sedimented by centrifugation at 7000 rev min^{-1} and 277 K for 20 min. They were resuspended in 1 l of the above medium without L-methionine. After another 1 h of cultivation at 303 K, 50 mg DL-selenomethionine was added. After an additional 30 min, the T7 promotor of the pET vector (Novagen) was induced with 1 m*M* IPTG. The cells were

Table 1
Summary of crystallographic data-collection and refinement statistics.

Values in parentheses refer to the highest resolution shell.

		SeMet MAD		
Data set	Native	Inflection	Peak	Remote
Data collection				
Wavelength (Å)	1.1271	0.97972	0.97997	0.98393
Resolution range (Å)	100–2.20 (2.28–2.20)	30–3.0 (3.11–3.00)	30–3.0 (3.11–3.00)	30–3.0 (3.11–3.00)
Space group	$I4_122$	$I4_122$	$I4_122$	$I4_122$
Unit-cell parameters				
a (Å)	112.2	113.2	113.2	113.2
b (Å)	112.2	113.2	113.2	113.2
c (Å)	335.5	335.9	335.9	335.9
Unique reflections	50618 (3259)	41207 (3982)	41235 (4004)	41270 (4079)
Completeness (%)	92.4 (60.7)	98.3 (98.6)	98.3 (98.5)	98.2 (98.3)
R_{sym}†	7.6 (36.9)	5.4 (25.1)	6.5 (32.6)	6.1 (27.5)
Average $I/\sigma(I)$	19.1 (1.4)	16.2 (4.3)	14.0 (3.4)	15.0 (4.1)
Multiplicity	11 (5)	4.3 (4.3)	4.3 (4.3)	4.3 (4.3)
Mosaicity (°)	0.33	0.39	0.39	0.35
MAD phasing				
Resolution range (Å)			30–3.5	
Overall figure of merit				
Before solvent flattening			0.62	
After solvent flattening			0.76	
Refinement statistics				
R_{cryst}/R_{free}‡ (%)	22.1/25.7			
Coordinate error§ (Å)	5.538			
No. of protein atoms	4346			
No. of ligand atoms	0			
No. of water molecules	317			
Ramachandran plot¶				
Most favourable regions (%)	90.5			
Additionally allowed regions (%)	9.2			
Generously allowed regions (%)	0.0			
Disallowed regions (%)	0.2			
R.m.s. deviations from ideality				
Bonds (Å)	1.697			
Angles (°)	0.017			
Average B values (Å^2)	60.3			
Protein residues/waters	518/317			

† $R_{sym} = 100 \times \sum_h \sum_i |I_i(h) - \langle I(h)\rangle| / \sum_h I(h)$, where $I_i(h)$ is the ith measurement of reflection h and $\langle I(h)\rangle$ is the average value of the reflection intensity. ‡ $R_{cryst} = \sum \big||F_o| - |F_c|\big| / \sum |F_o|$, where F_o and F_c are the structure-factor amplitudes from the data and the model, respectively. R_{free} is R_{cryst} using a 5% test set of structure factors. § Based on maximum likelihood. ¶ Calculated using *PROCHECK* (Laskowski *et al.*, 1993).

harvested following overnight growth by centrifugation at 6000*g* and 277 K for 20 min, washed in 20 m*M* Tris pH 8.0 and finally resuspended in 6 ml of the same buffer.

2.3. Purification and activity assay

The cell pellets (wet weight 5 and 1.6 g l^{-1} for wild-type AmyC and selenomethionine-labelled cells, respectively) were disrupted by two passages through a French pressure cell (American Instruments, Silver Springs, MD, USA). After separation from the cell debris by centrifugation (20 000*g*, 30 min, 277 K), the supernatants were incubated at 348 K for 20 min in order to denature the thermolabile host proteins, which were sedimented afterwards at 20 000*g*, 15 min, 277 K. The cleared supernatant, containing ∼3.4 mg protein, was dialyzed against 20 m*M* Tris–HCl pH 8.0 buffer and subjected to anion-exchange chromatography on a Source 30 Q HR10/10 column (Amersham Biosciences, Freiburg, Germany) equilibrated with the same buffer. Elution was performed using a linear gradient of NaCl (0–1 *M* in ten column volumes) in 20 m*M* Tris–HCl pH 8.0 buffer at a flow rate of 2 ml min^{-1}. Fractions containing AmyC, which eluted around 0.44 *M* NaCl, were pooled and dialyzed against 20 m*M* Tris–HCl pH 8.0 buffer. The total yield of purified protein was 1.5 mg. All buffers contained 5 m*M* DTT to prevent oxidation of DL-selenomethionine.

The activity assay was performed as described in Ballschmiter *et al.* (2006).

2.4. Crystallization and data collection

The protein solutions obtained were concentrated to 7.5 mg ml^{-1} using Ultra centrifugal filter devices (Millipore, Eschborn, Germany). Crystals were grown at 293 K using the sitting-drop vapour-diffusion method by mixing equal volumes of AmyC (14 mg ml^{-1} native and 8 mg ml^{-1} SeMet derivative) with reservoir solution (4 *M* sodium formate, 5% 2-propanol, 2 m*M* DTT). A three-wavelength MAD data set (high-energy remote, peak and inflection) was collected to 3.0 Å resolution using a flash-cooled SeMet AmyC crystal (4 *M* sodium formate, 5% 2-propanol, 2 m*M* DTT) at the PSF beamline BL2 at the Berliner Elektronen-Speicherring-Gesellschaft für Synchrotronstrahlung mbH (BESSY, Berlin, Germany). Typical data sets contained more than 100 frames with 1° rotation and 20 s exposure time per image. Native data were collected from a flash-cooled crystal (mother liquor containing 4 *M* sodium formate, 5% 2-propanol, 2 m*M* DTT) at PSF beamline BL1. The native crystal diffracted to a maximum resolution of 2.2 Å, with a mosaicity of 0.38°. The data were processed with *DENZO*/*SCALEPACK* (Table 1; HKL Research, Charlottesville, VA, USA) and revealed a centred tetragonal lattice with unit-cell parameters a = 112.2, b = 112.2, c = 335.6 Å, $\alpha = \beta = \gamma = 90°$. Systematic absences along the z axis revealed the space group to be $I4_122$. The Matthews coefficient (V_M = 4.3 Å^3 Da^{-1}) suggested the presence of one molecule in the asymmetric unit and a corresponding solvent content of 71%. X-ray data statistics are summarized in Table 1.

2.5. Structure determination and refinement

Data sets taken at the inflection point, peak and high-energy remote were used to obtain an initial electron-density map. A partial substructure (19 out of 21 Se sites) was determined with *SOLVE* (Terwilliger, 1996) using data to a

maximum resolution of 3.5 Å. The initial phases from *SOLVE* were improved by density modification using *RESOLVE* (Terwilliger, 2000), which resulted in overall figures of merit (FOMs) of 0.62 and 0.76, respectively (Table 1). A partial model (306 of 528 residues, 192 of which were built as alanines) was automatically built using autobuild as implemented in *RESOLVE* (Terwilliger, 2000) and was completed by manual fitting into σ_A-weighted $2|F_o - F_c|$ and $|F_o - F_c|$ difference electron-density maps (Read & Moult, 1992) using the programs *XTALVIEW* (McRee, 1999) and *O* (Jones *et al.*, 1991). The model was refined against the native data set (maximum resolution 2.2 Å) with *REFMAC*5 (Winn *et al.*, 2001) using standard parameters. A random set of 5% of reflections was excluded from refinement for cross-validation of various refinement strategies such as geometric and temperature-factor restrained values, the insertion of solvent water and as a basis for maximum-likelihood refinement using the *REFMAC*5 program and to monitor R_{free} (Brünger, 1993; Kleywegt & Brünger, 1996). Water molecules were assigned automatically for peaks $>2\sigma$ in $F_o - F_c$ difference maps by cycling the *REFMAC*5 refinement with *ARP*/*wARP* (Lamzin, 1993) and retained if they obeyed hydrogen-bonding criteria according to *HBPLUS* (McDonald & Thornton, 1994) and returned $2F_o - F_c$ density of $>1\sigma$ after refinement. The final model consists of one molecule (*X*) encompassing 528 residues. Owing to missing electron density, residues 405–414 are lacking and the side chains of the following residues were modelled as glycines: Phe466, Ile467, Thr469, Arg471, Thr472, Ser473, Val474 and Gln475. Alternate side-chain conformations were modelled for Glu90, Trp182 and Met227. The refinement statistics are summarized in Table 1. Surface-complementarity coefficients and solvent-accessible surface areas were calculated with *SC* using a 1.4 Å radius probe (Collaborative Computational Project, Number 4, 1994). Possible hydrogen bonds, salt bridges and van der Waals contacts were detected with *HBPLUS* and *CONTACSYM* (Sheriff *et al.*, 1987) using default parameters. Surface potentials were calculated with *PyMOL* using the implemented vacuum electrostatics function. The quality of the model was checked using *PROCHECK* (Laskowski *et al.*, 1993) and *WHATCHECK* (Hooft *et al.*, 1996). The Ramachandran plot (Ramachandran *et al.*, 1963) showed that 90.3% of the non-glycine and non-proline residues are in the most favourable region and 9.5% are in the additionally allowed region, while 0.0% are in the generously allowed region. Only Thr295 has a φ/ψ combination that lies in the disallowed regions of the Ramachandran plot, but the residue is very well defined in the electron-density map. This unusual arrangement is a consequence of the interactions of neighbouring residues, namely the main-chain N atom of Ser296 and the carbonyl O atom of Ile294, which form a hydrogen bond (distance 2.9 Å), thereby creating the unusual strain on Thr295. Further, the carbonyl O atom of Ile294 forms a bond with the ε-amino group of Lys304 (distance 3.0 A).

Secondary-structure elements were assigned using *STRIDE* (Eisenhaber *et al.*, 1995; Frishman & Argos, 1995). Coordinates were superimposed with *LSQKAB* (Kabsch *et al.*, 1976) from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994) or *LSQMAN* (Kleywegt, 1996). Figures were generated using *PyMOL* (http://www.pymol.org).

3. Results and discussion

3.1. Initial biochemical characterization and crystallization

The gene of a novel putative amylase has recently been identified in the genome of *T. maritima* (Ballschmiter *et al.*, 2006). In order to confirm that the gene indeed encodes an amylase, an initial characterization has been performed investigating the kinetics of the hydrolysis of starch and the influence of a potent amylase inhibitor, acarbose, on the activity of AmyC (Fig. 1). The Lineweaver–Burk plot (Fig. 1*a*)

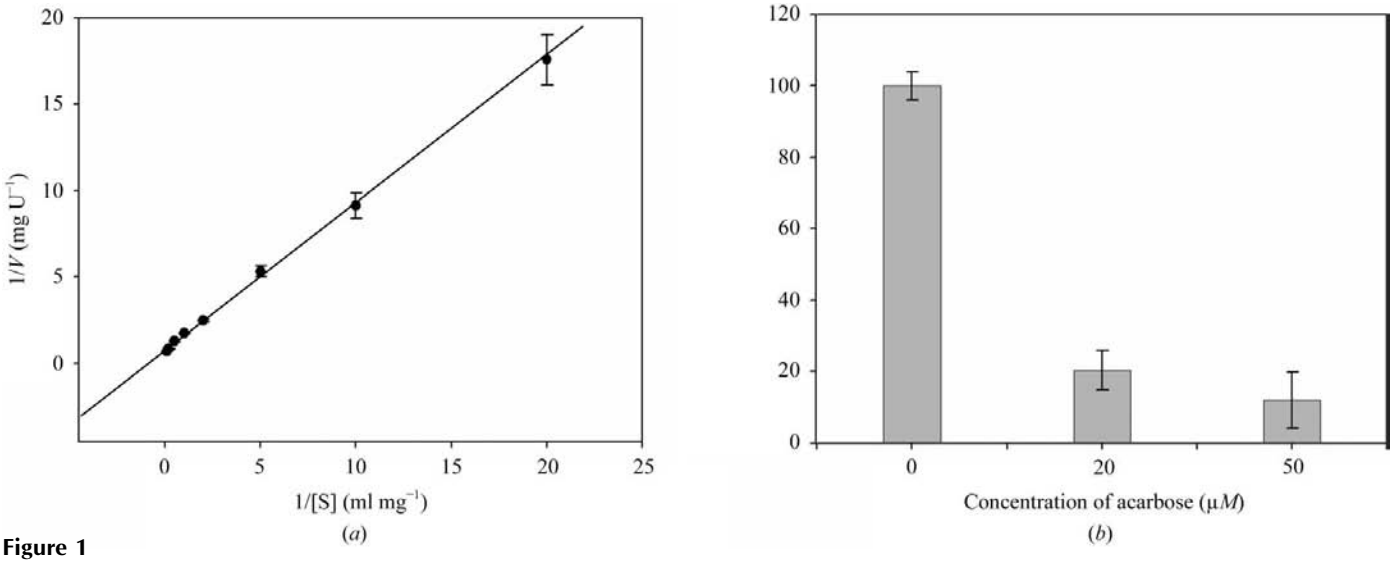

Figure 1
AmyC shows the characteristics of a functional amylase. (*a*) A Lineweaver–Burk diagram showing the kinetics of the hydrolysis of starch by AmyC allows the calculation of a K_m value of 1.1 mg ml^{-1} and a V_{max} of 1.3 U mg^{-1}. (*b*) AmyC activity is inhibited by the classical amylase inhibitor acarbose. The activity of Amy C is reduced to 25 or 20% using a 31-fold or 77-fold excess of acarbose, respectively.

allows the calculation of the K_m value for AmyC, which is 1.1 mg ml^{-1}, and V_{max}, which is 1.3 U mg^{-1}. The inhibitor acarbose strongly reduces the function of AmyC (Fig. 1*b*). These results and further experiments (Ballschmiter *et al.*, 2006) clearly demonstrate that AmyC is an amylase.

AmyC crystals were grown at 293 K using the sitting-drop vapour-diffusion method by mixing equal volumes of AmyC with reservoir solution comprising 4 *M* sodium formate, 5% 2-propanol, 2 m*M* DTT. The AmyC crystals obtained belong to the orthorhombic space group $I4_122$, with unit-cell parameters a = 112.2, b = 112.2, c = 335.5 Å. The Matthews coefficient of 4.1 Å^3 Da^{-1} corresponds to 70% solvent content and one molecule of AmyC in the asymmetric unit.

3.2. Structure determination and overall fold

The crystal structure of AmyC was determined by means of MAD using selenomethionine-containing AmyC (Table 1). The initial electron-density map obtained from experimental phases after solvent flattening was of high quality and allowed autotracing of 306 of the 528 amino acids present in the polypeptide chain. A native data set was used for refinement to a resolution of 2.2 Å (see §2; Table 1).

The model of AmyC encompasses all N- and C-terminal residues but lacks residues 404–414, which showed no interpretable electron density, indicating high flexibility of these residues. The structure was refined to a free R value of 27.7%, exhibiting good stereochemistry (Table 1). The overall structure of the molecule exhibits an almost perfect triangular shape that can be easily separated into two major parts, one before and one after the missing residues 405–414 (Fig. 2). The latter part (Fig. 2, coloured in green), forming the C-terminal domain, is composed of four α-helices forming a tightly packed bundle that is itself slightly supercoiled. This domain can be referred to as domain C with respect to the general overall structure of α-amylases. A closer look at the N-terminal domain reveals the other two conserved domains found in α-amylases: domain A, a TIM barrel (Fig. 2*a*, coloured red), and domain B, which is inserted between β-sheet 2 and α-helix 5 (coloured in blue). Domain A loosely resembles the classical $(\beta/\alpha)_8$-fold of α-amylases, with three structural features significantly differing from the canonical motif. Firstly, the barrel is made up of only seven parallel β-strands. Additionally, β-strands 4, 5 and 7 are quite short (two or three residues in length). Secondly, an additional β-strand (9) is indirectly involved in the formation of the barrel and does so by elongating β-strand 3 of the actual barrel at the N-terminal end (Figs. 2*a* and 2*b*). The obvious partitioning of the barrel into two distorted β-sheets (4–7 and 1–3, 8, 9) suggests that this might be mimicking the classical TIM-barrel motif (Fig. 2*a*).

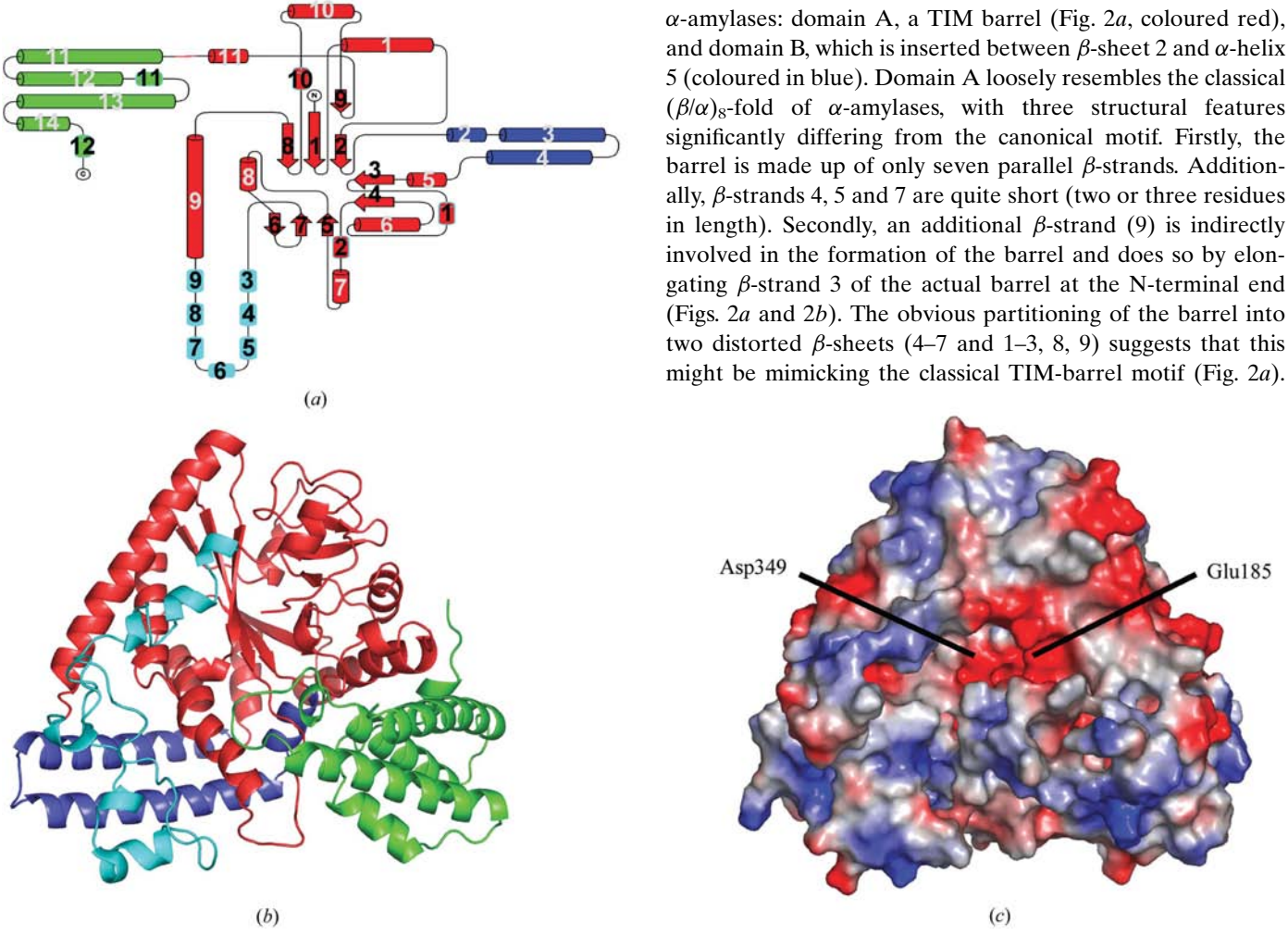

Figure 2
The AmyC monomer is made up of three domains. (*a*) The topology diagram of AmyC reveals a relationship to other α-amylases of GH-13, GH-70 and GH-77. β-Strands are represented by arrows and α-helices are shown as cylinders surrounded by black boxes, whereas 3_{10}-helices are shown as cylinders surrounded by cyan boxes. The AmyC monomer is coloured according to the common domain structure of α-amylases, with domain A depicted in red, domain B in blue and the region interacting with domain B in cyan. The C-terminal domain C is shown in green. Secondary-structure motifs (η denotes a 3_{10}-helix) are numbered consecutively from the N-terminus to the C-terminus. (*b*) Ribbon diagram of AmyC. The ribbon model of the AmyC monomer is also coloured according to the common domain structure of α-amylases, as in (*a*). (*c*) Surface potential of the side of AmyC bearing the putative catalytic residues. Areas coloured white, red and blue denote neutral, negative and positive potential, respectively.

Thirdly, the barrel is strongly flattened, resulting in a 'U' when viewed from the top. The gap is bridged by a 'lid' formed by

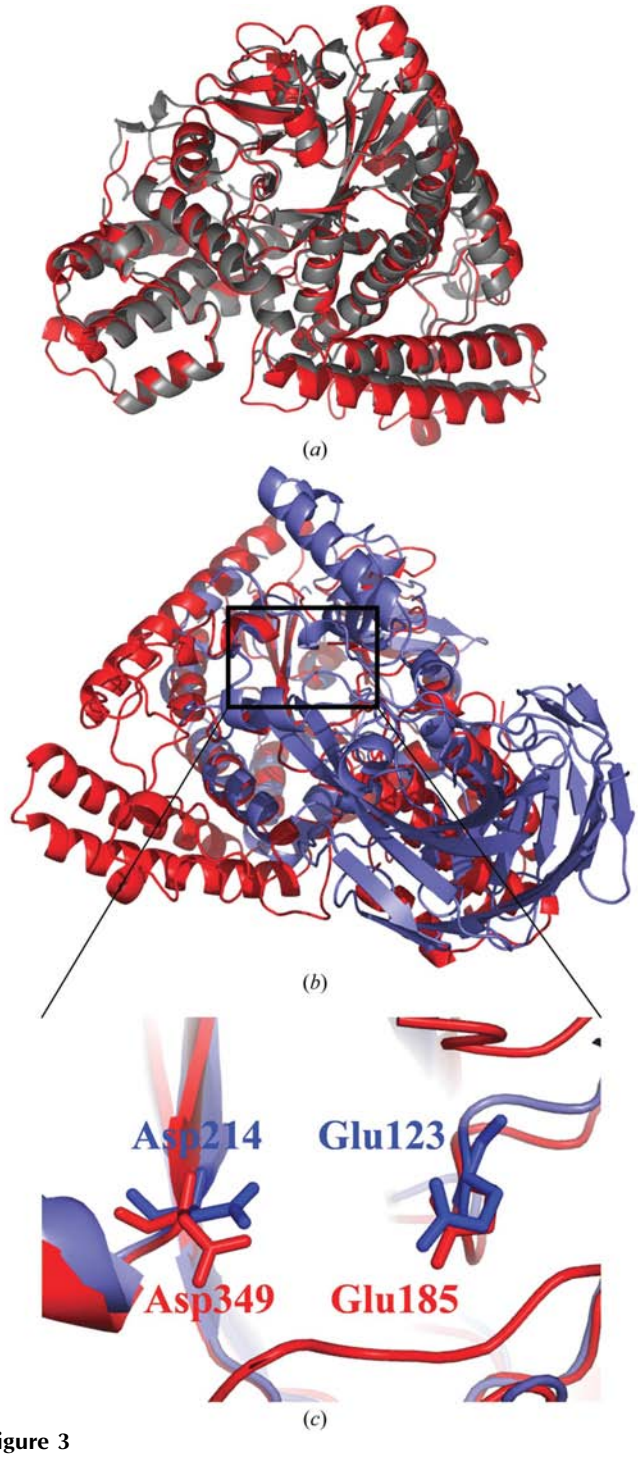

Figure 3
Structural comparison with other known structures of similar proteins. (*a*) Superposition of AmyC (red) with a putative amylase from *T. thermophilus* (PDB code 1ufa, grey), revealing an almost perfect structural identity. (*b*) The high structural similarity of AmyC (red) superpositioned with 4-α-glucano transferase (TLGT) from *Thermotoga litoralis* (PDB code 1k1w, slate) is mainly restricted to domain A and parts of domain C. (*c*) Three-dimensional enlargement of the part of domain A indicated by the black box in (*b*), representing the part of the TIM-barrel motif carrying the active residues. AmyC is shown in red and TLGT in slate, with the known catalytic residues of TLGT and the putative active residues of AmyC shown in a stick representation.

α-helix 8 and the residues following helix 11 as well as the antiparallel β-strand 6 (Figs. 2*a* and 2*b*).

According to the structural classification of proteins (SCOP; http://scop.berkeley.edu), this supersecondary structure of a seven-stranded β/α barrel [$(\beta/\alpha)_7$] fold has only been described and annotated in 11 proteins to date (Andreeva *et al.*, 2004; Murzin *et al.*, 1995), of which all except one, a putative phosphoesterase domain from *E. coli* (Teplyakov *et al.*, 2003), are polysaccharide-degrading or modifying enzymes. In addition to the 4-α-glucanotransferase from *T. litoralis* (Imamura *et al.*, 2003) belonging to GH-57, there are four cellulases and a putative polysaccharide deacetylase (PdaA) from *Bacillus subtilis* (Davies *et al.*, 2000; Spezio *et al.*, 1993; Varrot *et al.*, 2002, 2003; Zou *et al.*, 1999; Forouhar *et al.*, 2006). Additionally, two mannosidases from cow (*Bos taurus*; Heikinheimo *et al.*, 2003) and fruitfly (*Drosophila melanogaster*; van den Elsen *et al.*, 2001) belonging to GH-38 and a hyaluronidase from honey bee (*Apis mellifera*; Markovic Housley *et al.*, 2000) belonging to GH-56 have been shown to contain this motif. Sequential comparison of the putative members of glycosyl hydrolases in GH-38, which comprises 136 proteins, and GH-56, which contains 48 proteins, revealed that owing to sequence homology it can be speculated that all of them harbour the $(\beta/\alpha)_7$ fold.

Of the remaining two domains, domain B is formed by three α-helices (helices 2–4) interrupting the TIM barrel between β-sheet 2 and α-helix 5 and protruding into the solvent. An additional insertion found between β-strand 7 and α-helix 9 is composed of seven 3_{10}-helices (Fig. 2*a*, cyan). Their arrangement suggests two functions: they take part in the formation of the bottom of the barrel (with respect to Fig. 2*b* and to the location of the active site, see below) and shield this side of domains A and B from the solvent. A long α-helix precedes β-strand 9 and interacts with β-sheet 1, thus closing the barrel-like structure (Fig. 2*b*). Domain C fills one corner of the triangular-shaped AmyC and is composed of five α-helices and a 3_{10}-helix, with four of the α-helices forming a slightly superhelical bundle.

3.3. Quaternary structure

The crystal packing reveals clusters of four AmyC molecules, suggesting the existence of a homotetramer. One pair of AmyC molecules is arranged in a 'back to-back' conformation, with the second dimer placed on top but slightly rotated around a common central axis. This dimer of dimers is tightly packed and separated by water channels to neighbouring molecules. However, there is only one molecule per asymmetric unit and the other three molecules within the observed homotetramer are defined by crystal symmetry operations, raising the question whether this tetramer is a consequence of crystallization. Contradictory to this assumption is the fact that the contact area between the partners of a dimer is 1937 Å^2, which is at the upper end of the range of contact areas (800–2000 Å^2) found in other dimeric proteins (Bahadur *et al.*, 2003; Chakrabarti & Janin, 2002; Jones & Thornton, 1996; Ponstingl *et al.*, 2000). This supports the idea of at least

the formation of a functional dimer, instead of the multimerization being purely related to crystal packing. In the tetramer, the total contact area between the two dimers is 2214 Å^2, indicating the presence of a functional tetramer. This observation is supported by gel-filtration experiments performed by Ballschmiter *et al.* (2006), suggesting that the quaternary structure of AmyC consists of four molecules.

3.4. Structural comparison

A search in the *DALI* database (Holm & Sander, 1993) of families of structurally similar proteins using the full structure as a search model revealed significant homology to other known protein structures. A direct comparison of AmyC with itself gave a *Z* score of 63.9 and as expected an r.m.s.d. (root-mean-square deviation) of 0 Å. The *Z* score indicates the accuracy of the spatial localization of comparable C^{α} atoms and takes the number of similarly positioned residues into account. The highest *Z* score running *DALI* was achieved by comparison of AmyC with 1ufa, a protein of unknown function, with a *Z* score of 41.5 using 458 of the 501 residues. The 1ufa structure was solved as part of a structural genomics project on *T. thermophilus* Hb8 (Idaka *et al.*, 2006). 1ufa shows 35% sequence identity to AmyC and the superposition reveals an almost perfect overlap in all three domains (Fig. 3*a*), suggesting that the proteins AmyC and 1ufa serve identical cellular purposes (see below). So, in addition to AmyC, 1ufa can also be regarded as a new member of the SCOP class of $(\beta/\alpha)_7$-barrel proteins. A second protein with a *Z* score of 10.5, 1ny1 from *B. subtilis*, also a protein of unknown function, is assumed to be a polysaccharide hydrolase (Forouhar *et al.*, 2006) and exhibits a similar conformation in 182 out of its 231 residues with an r.m.s.d. of 3.2 Å. The third hit with a *Z* score above 10 was a glucanotransferase from *Thermotoga litoralis* (TLGT; PDB code 1k1w; Imamura *et al.*, 2003), which showed a *Z* score of 25.2 and an r.m.s.d. of 3.2 Å aligning 299 residues of its 615 residues to AmyC (14.1% identity between AmyC and TLGT). TLGT has been grouped into GH-57 owing to sequence similarities and its function as a glucanotransferase, which is well understood (Imamura *et al.*, 2001; Jeon *et al.*, 1997). The superposition of the two molecules (Fig. 3*b*) shows that the high similarity is mainly restricted to domain A (Fig. 3*c*), the TIM-barrel-like structure forming the active centre of the two enzymes, suggesting that domains B and C serve different purposes. This prompted us to run independent *DALI* searches for only domains B and C. As expected, domain B with a length of 56 residues as a three-helix domain revealed a *Z* score of 5.5 to 1ufa with an r.m.s.d. as low as 0.9 Å, although with a lowered sequence identity of 20% compared with the 35% for full-length 1ufa. Two more proteins, one involved in transcriptional regulation (PDB code1nka) and one in telomeric binding (PDB code 1k6o), revealed a low r.m.s.d. of 1.5 Å and a *Z* score comparable to that for 1ufa (6.0 and 5.9, respectively), but only 4 and 6% sequence identity, respectively. This and the fact that most of the other found hits with a *Z* score above 4 (data not shown) are involved in protein–protein interaction suggests that this domain B serves similar purposes in AmyC as well as in 1ufa.

Similar results were obtained for domain C. Domain C exhibits highest similarity to 1ufa (*Z* score 13.6 and r.m.s.d. of 2.1 Å). Interestingly, most of the other hits, which had a *Z* score of 7.5 or less and an r.m.s.d. in the range 2.2–4.1 Å, turned out to belong to hydrolases or oxidoreductases (data not shown). Surprisingly, the TLGT C-terminal domain also exhibits a significant structural similarity to AmyC in its C-terminal domain, but the *Z* score is quite low. The first three α-helices superpose well, but are decreased in length in TLGT. This helix bundle forms a bulge at one end of a groove containing the active site of TLGT (Fig. 3*b*). The fourth helix present in AmyC is missing in TLGT; instead, it carries a long C-terminal end forming an extended β-sheet structure. This overall structural similarity suggests that the mode of interaction for AmyC and 1ufa with polysaccharides is similar to that of TLGT.

Taken together, the overall structure of AmyC represents, after 1ufa, the second member of the COG1543 proteins to be crystallized. Both proteins exhibit a so far rare variation of the classical $(\beta/\alpha)_8$ barrel: the $(\beta/\alpha)_7$ barrel. The function of the COG1543 family is most likely to be the same or closely related to that of the GH-57 family. Interestingly, the protein family DUF200 ('domain of unknown function' in a different database, but strongly related to COG1543) has recently been merged with GH-57 owing to the high degree of similarity (see also http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF02651; Bateman *et al.*, 2004).

3.5. Active site

The overall structures of AmyC and 1ufa revealed a high degree of identity to TLGT in the organization and arrangement of the $(\beta/\alpha)_7$-barrel. The active site of TLGT has been characterized (Imamura *et al.*, 2001, 2003) and is located in the centre of the groove formed by the $(\beta/\alpha)_7$-barrel and the C-terminal helix bundle, with Glu123 and Asp214, which are located at the ends of β-strands 4 and 8, respectively, as the reactive amino acids (Fig. 3*c*). Sequence-based alignments have indicated that the active sites of enzymes belonging to COG1543 carry the same amino acids at these residues (Ballschmiter *et al.*, 2006). The structure comparison demonstrates that in AmyC, as well as in 1ufa, these sites are indeed occupied by the same amino acids, namely Glu (185 in AmyC, 184 in 1ufa) and Asp (349 in AmyC, 353 in 1ufa), respectively, pointing towards a function as catalytic residues. This is strengthened by the fact that in AmyC the surface area in the vicinity of the catalytic residues exhibits a charged (*i.e.* negative) surface potential as required for the binding of saccharide moieties (Fig. 2*c*). Polysaccharides such as amylase, starch and β-limit dextrin have been shown to be good substrates for AmyC (Ballschmiter *et al.*, 2006). Moreover, the location of two additional residues (His10 and His12) involved in polysaccharide binding close to the active site is highly conserved in AmyC, 1ufa, TLGT and COG1543. In TLGT the corresponding residues (His11 and His13) have been shown to

be involved in substrate binding (−1 subsite; Imamura *et al.*, 2003).

4. Conclusion

We have crystallized and solved the structure of the α-amylase AmyC from *T. maritima*, the first member of the COG1543 proteins with a known function (Ballschmiter *et al.*, 2006). The overall similarity within this family suggests that all members share a similar physiological function. A structure of a protein of unknown function has been deposited in the PDB (1ufa) which shows a very high degree of structural similarity to AmyC and its sequence is also highly similar to the members of COG1543. Thus, AmyC and 1ufa both harbour the uncommon $(\beta/\alpha)_7$-fold and could serve as prototype for this family of proteins. Furthermore, a 4-glucanotransferase from *T. litoralis* exhibits the same fold in the catalytically active domain A as AmyC. In addition, the fact that AmyC as well as the 4-glucanotransferase display a similar function and arrangement of the active site suggests that the members of COG1543 form a subgroup within GH-57, sharing the $(\beta/\alpha)_7$-fold.

We thank the staff of PSF Beamline BL1 at BESSY (Berlin) for excellent support during data collection. This work was supported by the Deutsche Forschungs-Gemeinschaft to WL (Li 398/6).

References

Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J., Chothia, C. & Murzin, A. G. (2004). *Nucleic Acids Res.* **32**, D226–D229.
Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. (2003). *Proteins*, **53**, 708–719.
Ballschmiter, M., Fütterer, O. & Liebl, W. (2006). Submitted.
Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). *Nucleic Acids Res.* **32**, D138–D141.
Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
Chakrabarti, P. & Janin, J. (2002). *Proteins*, **47**, 334–343.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Davies, G. J., Brzozowski, A. M., Dauter, M., Varrot, A. & Schulein, M. (2000). *Biochem. J.* **348**, 201–207.
Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem. J.* **321**, 557–559.
Eisenhaber, F., Persson, B. & Argos, P. (1995). *Crit. Rev. Biochem. Mol. Biol.* **30**, 1–94.
Elsen, J. M. van den, Kuntz, D. A. & Rose, D. R. (2001). *EMBO J.* **20**, 3008–3017.
Forouhar, F., Edstrom, W., Khan, J., Ma, L., Chiang, Y., Acton, T. B., Montelione, G., Hunt, J. F. & Tong, L. (2006). In preparation.
Frishman, D. & Argos, P. (1995). *Proteins*, **23**, 566–579.
Heikinheimo, P., Helland, R., Leiros, H. K., Leiros, I., Karlsen, S., Evjen, G., Ravelli, R., Schoehn, G., Ruigrok, R., Tollersrud, O. K., McSweeney, S. & Hough, E. (2003). *J. Mol. Biol.* **327**, 631–644.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.
Horvathova, V., Janecek, S. & Sturdik, E. (2001). *Gen. Physiol. Biophys.* **20**, 7–32.
Idaka, M., Terada, T., Murayama, K., Yamaguchi, H., Nureki, O., Ishitani, R., Kuramitsu, S., Shirouzu, M. & Yokoyama, S. (2006). In preparation.
Imamura, H., Fushinobu, S., Jeon, B. S., Wakagi, T. & Matsuzawa, H. (2001). *Biochemistry*, **40**, 12400–12406.
Imamura, H., Fushinobu, S., Yamamoto, M., Kumasaka, T., Jeon, B. S., Wakagi, T. & Matsuzawa, H. (2003). *J. Biol. Chem.* **278**, 19378–19386.
Imamura, H., Jeon, B., Wakagi, T. & Matsuzawa, H. (1999). *FEBS Lett.* **457**, 393–396.
Jeon, B. S., Taguchi, H., Sakai, H., Ohshima, T., Wakagi, T. & Matsuzawa, H. (1997). *Eur. J. Biochem.* **248**, 171–178.
Jones, S. & Thornton, J. M. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W., Kabsch, H. & Eisenberg, D. (1976). *J. Mol. Biol.* **100**, 283–291.
Kleywegt, G. J. (1996). *Acta Cryst.* D**52**, 842–857.
Kleywegt, G. J. & Brünger, A. T. (1996). *Structure*, **4**, 897–904.
Lamzin, V. S. (1993). *Acta Cryst.* D**49**, 129–147.
Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.
Liebl, W., Stemplinger, I. & Ruile, P. (1997). *J. Bacteriol.* **179**, 941–948.
Lim, W. J., Park, S. R., An, C. L., Lee, J. Y., Hong, S. Y., Shin, E. C., Kim, E. J., Kim, J. O., Kim, H. & Yun, H. D. (2003). *Res. Microbiol.* **154**, 681–687.
McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Markovic Housley, Z., Miglierini, G., Soldatova, L., Rizkallah, P. J., Muller, U. & Schirmer, T. (2000). *Structure Fold. Des.* **8**, 1025–1035.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Nelson, K. E. *et al.* (1999). *Nature (London)*, **399**, 323–329.
Nielsen, J. E. & Borchert, T. V. (2000). *Biochim. Biophys. Acta*, **1543**, 253–274.
Ponstingl, H., Henrick, K. & Thornton, J. M. (2000). *Proteins*, **41**, 47–57.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Read, R. J. & Moult, J. (1992). *Acta Cryst.* A**48**, 104–113.
Reuter, K., Nottrott, S., Fabrizio, P., Luhrmann, R. & Ficner, R. (1999). *J. Mol. Biol.* **294**, 515–525.
Sheriff, S., Hendrickson, W. A. & Smith, J. L. (1987). *J. Mol. Biol.* **197**, 273–296.
Spezio, M., Wilson, D. B. & Karplus, P. A. (1993). *Biochemistry*, **32**, 9906–9916.
Svensson, B. (1994). *Plant Mol. Biol.* **25**, 141–157.
Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997). *Science*, **278**, 631–637.
Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J. & Natale, D. A. (2003). *BMC Bioinformatics*, **11**, 41–55.
Teplyakov, A., Obmolova, G., Khil, P. P., Howard, A. J., Camerini-Otero, R. D. & Gilliland, G. L. (2003). *Proteins*, **51**, 315–8.
Terwilliger, T. C. (1996). *Acta Cryst.* D**52**, 749–757.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Varrot, A., Frandsen, T. P., Driguez, H. & Davies, G. J. (2002). *Acta Cryst.* D**58**, 2201–2204.
Varrot, A., Frandsen, T. P., von Ossowski, I., Boyer, V., Cottaz, S., Driguez, H., Schulein, M. & Davies, G. J. (2003). *Structure*, **11**, 855–864.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Zechel, D. L., He, S., Dupont, C. & Withers, S. G. (1998). *Biochem. J.* **336**, 139–145.

Zechel, D. L. & Withers, S. G. (2000). *Acc. Chem. Res.* **33**, 11–18.
Zona, R., Chang-Pi-Hin, F., O'Donohue, M. J. & Janecek, S. (2004). *Eur. J. Biochem.* **271**, 2863–2872.
Zou, J., Kleywegt, G. J., Stahlberg, J., Driguez, H., Nerinckx, W., Claeyssens, M., Koivula, A., Teeri, T. T. & Jones, T. A. (1999). *Structure Fold. Des.* **7**, 1035–1045.