Three-dimensional structure of a variant 'Termamyl-like' *Geobacillus stearothermophilus* α -amylase at 1.9 Å resolution

Wendy A. Offen,^a Anders Viksoe-Nielsen,^b Torben V. Borchert,^b Keith S. Wilson^a and Gideon J. Davies^{a*}

^aYork Structural Biology Laboratory, Department of Chemistry, The University of York, York YO10 5DD, England, and ^bNovozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark

Correspondence e-mail: gideon.davies@york.ac.uk

Received 2 October 2014
Accepted 2 December 2014

PDB reference: engineered bacterial α -amylase, 4uзу

The enzyme-catalysed degradation of starch is central to many industrial processes, including sugar manufacture and first-generation biofuels. Classical biotechnological platforms involve steam explosion of starch followed by the action of endo-acting glycoside hydrolases termed α -amylases and then exo-acting α -glucosidases (glucoamylases) to yield glucose, which is subsequently processed. A key enzymatic player in this pipeline is the 'Termamyl' class of bacterial α -amylases and designed/evolved variants thereof. Here, the three-dimensional structure of one such Termamyl α -amylase variant based upon the parent *Geobacillus stearothermophilus* α -amylase is presented. The structure has been solved at 1.9 Å resolution, revealing the classical three-domain fold stabilized by Ca^{2+} and a Ca^{2+} – Na^{+} – Ca^{2+} triad. As expected, the structure is similar to the *G. stearothermophilus* α -amylase but with main-chain deviations of up to 3 Å in some regions, reflecting both the mutations and differing crystal-packing environments.

1. Introduction

The world starch market reflects the utilization of approximately 75 million tons of starch per annum. α -Amylases (EC 3.2.1.1), which are endo-acting starch-degrading glycoside hydrolases, are therefore key industrial enzymes of the modern age. There is massive application of α -amylases in diverse biotechnological processes, including the production of high-fructose corn syrups and the conversion of corn starch to a first-generation biofuel, as well as in the detergent, pharmaceutical and baking industries.

The majority of the commercially relevant α -amylases are found in family GH13 of the CAZy (<http://www.cazy.org>; recently reviewed in Cantarel *et al.*, 2009) sequence-based classification of glycoside hydrolases (for a historical context, see Henrissat & Davies, 1997). CAZy family GH13 is one of the better studied families of glycoside hydrolases. Even at the structural level there are, at the time of submission, almost 400 PDB entries reflecting almost 100 three-dimensional structures of different enzymes (best reviewed at http://www.cazy.org/GH13_structure.html).

The 'Termamyl' subgrouping of GH13 α -amylases, named after an industrially useful enzyme from *Bacillus licheniformis*, has particular industrial application. These enzymes typically display a core three-domain (termed A, B and C) structure. The substrate-binding cleft and hence the catalytic centre lie at the interface of the A and B domains. As first revealed in the three-dimensional structures of the *B. licheniformis* enzyme (Machius *et al.*, 1998) and a hybrid *B. licheniformis*/*B. amyloliquefaciens* enzyme (Brzozowski *et al.*, 2000), these enzymes display an unusual Ca^{2+} – Na^{+} – Ca^{2+} triad at the A/B-domain interface as well as displaying differing numbers of Ca^{2+} ions in their C-terminal domains. Three-dimensional structures of these α -amylases have proved to be extremely useful in the industrial context, not least because of the success of rational protein engineering of these systems (reviewed, for example, in Shaw *et al.*, 1999; Nielsen & Borchert, 2000). Currently, three-dimensional structures of the 'Termamyl-like' amylases include enzymes from the following *Bacillus* species: *B. licheniformis* (Machius *et al.*, 1995), a chimeric *B. licheniformis*/*B. amyloliquefaciens* enzyme (Brzozowski *et al.*, 2000), *Geobacillus stearothermophilus* (Suvd *et al.*, 2001), *B. halmapalus* (Davies *et al.*, 2005), *Bacillus* sp. 707 (Kanai *et al.*, 2004),

Table 1X-ray data collection and structure refinement statistics for the TVB146 α -amylase.

Values in parentheses are for the highest resolution shell.

Space group	<i>I</i> 4 ₁ 22
Unit-cell parameters (Å, °)	<i>a</i> = <i>b</i> = 108.18, <i>c</i> = 181.49, α , β , γ = 90
Data processing	
Resolution range (Å)	34.21–1.90 (2.00–1.90)
<i>R</i> _{merge}	0.120 (0.518)
$\langle I \rangle / \langle \sigma(I) \rangle$	17.6 (5.4)
Completeness (%)	100 (100)
Multiplicity	14.7 (14.8)
Model refinement	
No. of reflections used	40518
<i>R</i> _{cryst} / <i>R</i> _{free}	0.17/0.21
No. of protein atoms	3855
No. of ligand atoms	3 Ca ²⁺ , 3 Na ⁺ , 2 Cl [−] , 10 glycine
No. of water atoms	408
R.m.s.d., bonds (Å)	0.017
R.m.s.d., angles (°)	1.38
Mean <i>B</i> values (Å ²)	
Protein atoms	
Main chain	17
Side chain	17
Ligand atoms	32
Overall	32
Ca ²⁺	14
Na ⁺	24
Cl [−]	8
Glycine	47
Waters	27
Ramachandran plot† (%)	
Favoured	97.3
Allowed	2.7
PDB code	4uzu

† Calculated using the validation options in *Coot*.

Bacillus sp. KSM-1378 (Shirai *et al.*, 2007), *B. amyloliquefaciens* (Alikhajeh *et al.*, 2010) and *Bacillus* sp. KR-8104 (PDB entry 3dc0; J. Alikhajeh, K. Khajeh, B. Ranjbar, M. Naderi-Manesh, H. Naderi-Manesh & C.-J. Chen, unpublished work).

Here, we report the three-dimensional structure solution and refinement of a variant (hereafter termed TVB146), featuring both a two-residue deletion and a point substitution, of the *G. stearothermophilus* ‘Termamyl-like’ α -amylase. The structure has been refined at a resolution of approximately 1.9 Å, revealing a number of conformational changes relative to the known structure of the closest homologue, the majority of which occur adjacent to the structural truncation.

2. Materials and methods

2.1. Protein production

The amylase gene was expressed by heterologous expression in a *B. licheniformis* host. The amylase was purified by diluting a 10 ml protein sample with 290 ml buffer *A* (5 mM glycine pH 9.5, 2 mM CaCl₂). This was applied onto a Q-Sepharose column equilibrated in buffer *A* at a flow rate of 10 ml min^{−1}. The enzyme was eluted from the column using a 0–1 M NaCl gradient in buffer *A*; the protein eluted from the column at approximately 0.15 M NaCl. Fractions of 10 ml were collected and those containing active enzyme were combined and subsequently used for crystallization trials.

2.2. Crystallization and data collection

The protein was crystallized by the hanging-drop method at a concentration of 10 mg ml^{−1} in 3 mM glycine, 3 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 6.5, 0.15 M NaCl, 1 mM CaCl₂ with the addition of 1 mM of the pseudotetrasaccharide inhibitor

acarbose. The well solution consisted of 65%(v/v) 2-methyl-2,4-pentanediol (MPD), 0.1 M 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid (HEPES) pH 8.2, and the drop was comprised of 1 µl protein solution and 0.7 µl well solution. Crystals grew at 18°C over a period of three weeks. For data collection, additional cryoprotectant was not needed as the crystallization drop contained a sufficiently high concentration of MPD. The crystal was therefore harvested directly into liquid nitrogen using a nylon-fibre CryoLoop (Hampton Research, Aliso Viejo, California, USA). Data were collected at a wavelength of 0.8276 Å on beamline ID23-2 at the European Synchrotron Radiation Facility. All computational analysis used programs from the *CCP4* suite (Winn *et al.*, 2011) unless otherwise stated. The images were processed in *MOSFLM* and scaled in *SCALA*. The space group was *I*4₁22, with unit-cell parameters *a* = *b* = 108.2, *c* = 181.5 Å. Data processing and structure refinement details are given in Table 1.

2.3. Structure solution and refinement

The structure was solved by molecular replacement using *Phaser* (McCoy *et al.*, 2007) with PDB entry 1hvx as the model (α -amylase from *G. stearothermophilus*; 97% identity; Suvd *et al.*, 2001). Manual correction of the model and insertion of solvent were performed using *Coot* (Emsley *et al.*, 2010). Two additional sodium ions were modelled. One was recognized as a potential sodium ion from its coordination by three water molecules, Asp232 OD1 and Glu262 OE1 (in the active centre; discussed below), and the other was flagged as having five waters coordinated when validation checks were run in *Coot* (Emsley *et al.*, 2010); the additions were confirmed as suitable from the resultant *F_o − F_c* and 2*F_o − F_c* electron-density maps. The structure was refined using *REFMAC5* (Murshudov *et al.*, 2011). Surprisingly, acarbose was not observed in the three-dimensional structure for reasons that are unclear.

Coordinates have been deposited with the PDB with accession code 4uzu.

3. Results and discussion

3.1. Structure of the TVB146 amylase

TVB146 is a variant of the *G. stearothermophilus* (originally named *B. stearothermophilus*) ‘Termamyl-like’ α -amylase represented in PDB entry 1hvx (Suvd *et al.*, 2001). In comparison with the most similar deposited sequence, the TVB146 enzyme reflects a two-residue deletion (Δ 181–182; discussed below) and a single-point variant N193F (1hvx numbering; corresponding to Phe191 in 4uzu). In comparison with the sequence present in 1hvx, the TVB146 enzyme shows 97% sequence identity and differs, in addition to the two-residue deletion Δ 181–182, at A73T, N193F, S217N, M278T, N281D, T304A and V416G. There is no electron density observed for residues 484–515 in either 1hvx or for the corresponding residues in 4uzu (except for the main-chain and C ^{β} atoms of Lys482 in the latter).

The overall structure comprises a typical ‘Termamyl’ α -amylase three-domain assembly with domain A, a (β/α)₈-barrel, at the core (Fig. 1*a*). Domain B extends from between β -strand 3 and α -helix 3 of domain A, and the C-terminal domain C lies on the opposite side of the core. Domain B consists of three antiparallel β -sheets, one four-stranded sheet with a two-stranded sheet on a long loop extending between the first and third strands, and a short two-stranded sheet on a loop between the third and fourth strands. The Ca²⁺–Na⁺–Ca²⁺ binding site (Fig. 1*b*) is located within the B domain, extending towards the central domain. Domain C has a globular structure comprising an eight-stranded β -sheet in which four strands are folded

Figure 1

Three-dimensional structure of TVB146 amylase (PDB entry 4uzu). (a) Stereoview of a ribbon diagram of TVB146 amylase showing domains A (red), B (green) and C (blue). Spheres represent ions: calcium in tan, sodium in yellow and chloride in grey. (b) Observed $F_o - F_c$ OMIT density (contoured at $0.4 \text{ e} \text{ \AA}^{-3}$, equivalent to 3σ) for the region around the Ca^{2+} - Na^{+} - Ca^{2+} triad with waters shown as red spheres and hydrogen bonds shown as dashed lines. This figure was produced using *CCP4mg* (McNicholas *et al.*, 2011).

in a Greek-key motif. The third Ca^{2+} ion binds at the interface of domains C and A. These features are also present in 1hvx; in addition there are a further two sodium ions, two chloride ions (both at partial occupancy) and two glycine molecules from the crystallization buffer in 4uzu. The chloride ions are notable in that they are both involved in crystal packing. One chloride ion is positioned at the intersection

of three crystallographic twofold axes, whilst the other lies between two molecules along one of the twofold axes. The presence of the chloride ions may thus facilitate the crystallization of the TVB146 amylase in a tetragonal space group; 1hvx was obtained from a crystal grown in the monoclinic space group $P2_1$ under different conditions (Suvd *et al.*, 2001). A sodium ion also makes a bridge to a symmetry-related molecule by coordinating to water molecules that are hydrogen-bonded to each protein molecule (to Asp125 OD1 and OD2 and Asn127 ND2 on one molecule and to Glu129 OE1 and Ser131 O on the other). A second 'additional' Na^{+} ion is found bound in the active centre, as discussed below.

Figure 2

Graph showing C^α deviations between 4uzu and 1hvx, omitting loop Ile181–Gly182 of 1hvx (the numbering refers to 1hvx). Structures were overlaid with *PDBFold* and the residue-by-residue C^α deviation is shown.

3.2. Comparison with the 'parent' *G. stearothermophilus* α -amylase

The TVB146 amylase structure was compared with the closest three-dimensional structure, PDB entry 1hvx, with which it shares 97% sequence identity. Using *PDBFold* (Krissinel & Henrick, 2004) the two structures share a C^α r.m.s.d. of $0.56 \text{ \AA}$. A graph showing the residue-by-residue deviations between the C^α atoms of 4uzu and 1hvx is shown in Fig. 2. Unsurprisingly, the greatest deviations, with C^α movements up to $3.1 \text{ \AA}$, are observed flanking the loop deletion at

Figure 4

Active centre of TVB146 amylase showing how the binding of Na^+ mimics the charge developed during catalysis. (a) Simplified schematic diagram showing the transition state for glycosyl-enzyme formation of an α -amylase. (b) Overlay of residues Asp232, Glu262 and Asp329 from the *B. halmapalus* α -amylase (PDB entry 1w9x, in green; Davies *et al.*, 2005) with the equivalent residues Asp232, Glu262 and Asp329 from 4uzu (coral). The disaccharide portion of acarbose occupying the +1 and -1 subsites in 1w9x and the active-site sodium ion (in yellow) from 4uzu are also shown. Electron density for the sodium ion in an $F_o - F_c$ OMIT map contoured at $0.4 \text{ e } \text{\AA}^{-3}$ is shown. This figure was produced using CCP4mg.

Figure 3

Overlap of the key loop region of the TVB146 amylase (green) with that from *G. stearothermophilus* (PDB entry 1hvx) (dark purple). Differences in hydrogen bonding described in the text are shown, with hydrogen bonds represented as dashed lines annotated with distances in Å. Water molecules from 1hvx are shown as red spheres.

$\Delta 181-182$. In 1hvx the loop features two hydrogen bonds between the main-chain carbonyl O atom of Gly180 and the main-chain N atoms of Gly182 and Lys183 (Fig. 3). There is also a water molecule which is hydrogen-bonded between the main-chain carbonyl O atoms of Phe178 and Lys183, and a further water molecule interacting with this water and the N and carbonyl O atoms of Gly180. In 4uzu, without the loop, the distance is less and Phe178 O is able to hydrogen-bond directly to Lys181 N ($3.0 \text{ \AA}$).

The variation in the regions flanking the loop seems to have ramifications for the adjacent three-dimensional environment, as the other area of largest movement (approximately residues 117–134) is the two β -strands that form a sheet with 174–179, whose position is perturbed by the deletion itself. However, some differences in these regions may be also affected by interactions with symmetry-related molecules in 4uzu that are different from the interactions of these regions in 1hvx. For example, Asp117 OD2 is hydrogen-bonded to Trp185 NE1 in one symmetry-related molecule, whilst Asn122 ND2 is hydrogen-bonded to the main-chain O of Ile130 ($2.8 \text{ \AA}$), and Glu128 NE2 is hydrogen-bonded to Glu129 O ($2.9 \text{ \AA}$), in another symmetry-related molecule. Glu128 NE2 is also $3.3 \text{ \AA}$ from the chloride ion at the intersection of three crystallographic twofold axes. Thr133 OG is able to hydrogen-bond to Arg179 NE ($2.9 \text{ \AA}$) in 4uzu,

whereas in 1hvx it hydrogen-bonds the NH1 atom of the arginine ($3.2 \text{ \AA}$); thus, the guanidinium group in 1hvx is displaced further from Glu129, with which it forms one hydrogen bond ($2.7 \text{ \AA}$), whereas in 4uzu both Arg179 NH1 and NH2 are hydrogen-bonded to the carboxyl group of Glu129 (2.9 and $3.0 \text{ \AA}$). In addition, in 4uzu Lys177 NZ is $3.2 \text{ \AA}$ from the chloride ion on the crystallographic twofold axis. Whilst these bonding interactions may contribute to the differences in the positions of the strands, they may not reflect differences that exist for the enzymes in solution.

A further region in which significant differences in C^α position are observed (based solely on the global superposition) is around the Ca^{2+} ion found in domain C; the reasons for this are unclear as 4uzu and 1hvx are almost identical in sequence, and are essentially free from crystal-packing contacts, in this region. The point mutation N193F (1hvx numbering, corresponding to Phe191 in 4uzu) is positioned close to the α -glucan substrate-binding groove close to the +2 subsite (subsite nomenclature is given in Davies *et al.*, 1997).

4. Discussion

The TVB146 α -amylase is a classical 'Termamyl-like' α -amylase from CAZY family GH13. This is one of the best studied of all glycoside

hydrolase families. Mechanistically, the enzymes perform hydrolysis of the α -1,4-glycosidic bonds of starch with net retention of anomeric configuration *via* a covalent glycosyl-enzyme intermediate. The mechanism is well defined, with requirements for a catalytic nucleophile, a Brønsted acid/base and a 'helper' residue which stabilizes the transition state through coordination to the O2 and O3 hydroxyls of the -1 subsite sugar. In the case of the TVB146 amylase variant, homology with well characterized enzymes allows us to say that these three residues are Asp232, Glu262 and Asp329, respectively. It was surprising that this structure, which was co-crystallized in the presence of the inhibitor acarbose, showed no density for this species; the reasons for this are unclear. A sodium ion is found in the catalytic centre interacting with both the catalytic nucleophile Asp232 and the Brønsted acid/base Glu262. Notably, the positively charged ion binds almost exactly where the positive charge at the anomeric C atom would be in the oxocarbenium-ion-like transition state of the reaction (Fig. 4). Such binding of monovalent cations, mimicking the transition-state charge, has been observed previously for a number of glycoside hydrolases (for a recent example, see Thompson *et al.*, 2012).

Many α -amylase variants with improved stability reflect variants in the vicinity of the A/B-domain interface, which is believed to be less stable than other areas of the structure (as discussed, for example, in Declerck *et al.*, 2000; Machius *et al.*, 2003). A two-residue deletion in the *B. amyloliquefaciens* α -amylase was reported by Suzuki *et al.* (1989) and shown to be more stable. The TVB146 α -amylase is a designed industrial variant notably featuring a deletion of two residues that results in a much tighter turn, and less exposed loop, than in the parent enzyme. The TVB146 α -amylase structure thus adds to the growing knowledge of the structure–function relationships in stable industrially relevant α -amylases.

KSW acknowledges grant support from Novozymes A/S for aspects of this work. The staff of the European Synchrotron Radiation Facility are thanked for beamline provision.

References

- Alikhajeh, J., Khajeh, K., Ranjbar, B., Naderi-Manesh, H., Lin, Y.-H., Liu, E., Guan, H.-H., Hsieh, Y.-C., Chuankhayan, P., Huang, Y.-C., Jeyaraman, J., Liu, M.-Y. & Chen, C.-J. (2010). *Acta Cryst.* **F66**, 121–129.
- Brzozowski, A. M., Lawson, D. M., Turkenburg, J. P., Bisgaard-Frantzen, H., Svendsen, A., Borchert, T. V., Dauter, Z., Wilson, K. S. & Davies, G. J. (2000). *Biochemistry*, **39**, 9099–9107.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, d233–D238.
- Davies, G. J., Brzozowski, A. M., Dauter, Z., Rasmussen, M. D., Borchert, T. V. & Wilson, K. S. (2005). *Acta Cryst.* **D61**, 190–193.
- Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem. J.* **321**, 557–559.
- Declerck, N., Machius, M., Wiegand, G., Huber, R. & Gaillardin, C. (2000). *J. Mol. Biol.* **301**, 1041–1057.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Henrissat, B. & Davies, G. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.
- Kanai, R., Haga, K., Akiba, T., Yamane, K. & Harata, K. (2004). *Biochemistry*, **43**, 14047–14056.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Machius, M., Declerck, N., Huber, R. & Wiegand, G. (1998). *Structure*, **6**, 281–292.
- Machius, M., Declerck, N., Huber, R. & Wiegand, G. (2003). *J. Biol. Chem.* **278**, 11546–11553.
- Machius, M., Wiegand, G. & Huber, R. (1995). *J. Mol. Biol.* **246**, 545–559.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* **D67**, 386–394.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Nielsen, J. E. & Borchert, T. V. (2000). *Biochim. Biophys. Acta*, **1543**, 253–274.
- Shaw, A., Bott, R. & Day, A. G. (1999). *Curr. Opin. Biotechnol.* **10**, 349–352.
- Shirai, T., Igarashi, K., Ozawa, T., Hagihara, H., Kobayashi, T., Ozaki, K. & Ito, S. (2007). *Proteins*, **66**, 600–610.
- Suud, D., Fujimoto, Z., Takase, K., Matsumura, M. & Mizuno, H. (2001). *J. Biochem.* **129**, 461–468.
- Suzuki, Y., Ito, N., Yuuki, T., Yamagata, H. & Udaka, S. (1989). *J. Biol. Chem.* **264**, 18933–18938.
- Thompson, A. J., Heu, T., Shaghasi, T., Benyamino, R., Jones, A., Friis, E. P., Wilson, K. S. & Davies, G. J. (2012). *Acta Cryst.* **D68**, 875–882.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.