STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

# Structure and function of the type III pullulan hydrolase from *Thermococcus kodakarensis*

**Jingxu Guo, Alun R. Coker, Steve P. Wood, Jonathan B. Cooper, Ronan M. Keegan, Nasir Ahmad, Majida Atta Muhammad, Naeem Rashid and Muhummad Akhtar**

*Acta Cryst.* (2018). D**74**, 305–314

STRUCTURAL BIOLOGY

ISSN 2059-7983

# Structure and function of the type III pullulan hydrolase from *Thermococcus kodakarensis*

**Jingxu Guo,[a] Alun R. Coker,[a] Steve P. Wood,[a] Jonathan B. Cooper,[a,b]* Ronan M. Keegan,[c,d] Nasir Ahmad,[e] Majida Atta Muhammad,[f] Naeem Rashid[f] and Muhummad Akhtar[f]**

[a]Division of Medicine, University College London, Gower Street, London WC1E 6BT, England, [b]Department of Biological Sciences, Birkbeck, University of London, Malet Street, Bloomsbury, London WC1E 7HX, England, [c]CCP4, Research Complex at Harwell and Science and Technology Facilities Council, Rutherford Appleton Laboratories, Harwell Oxford, Didcot OX11 0FA, England, [d]Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, England, [e]Institute of Agricultural Sciences, University of the Punjab, Quaid-e-Azam Campus, Lahore 54590, Pakistan, and [f]School of Biological Sciences, University of the Punjab, Quaid-e-Azam Campus, Lahore 54590, Pakistan. *Correspondence e-mail: jon.cooper@ucl.ac.uk

Received 27 August 2017
Accepted 29 January 2018

Edited by Z. S. Derewenda, University of Virginia, USA

**Keywords:** pullulan hydrolase; *Thermococcus kodakarensis*; protein crystallography; structural biology.

**PDB reference**: type III pullulan hydrolase, 5ot1

**Supporting information**: this article has supporting information at journals.iucr.org/d

Pullulan-hydrolysing enzymes, more commonly known as debranching enzymes for starch and other polysaccharides, are of great interest and have been widely used in the starch-saccharification industry. Type III pullulan hydrolase from *Thermococcus kodakarensis* (TK-PUL) possesses both pullulanase and $\alpha$-amylase activities. Until now, only two enzymes in this class, which are capable of hydrolysing both $\alpha$-1,4- and $\alpha$-1,6-glycosidic bonds in pullulan to produce a mixture of maltose, panose and maltotriose, have been described. TK-PUL shows highest activity in the temperature range 95–100°C and has a pH optimum in the range 3.5–4.2. Its unique ability to hydrolyse maltotriose into maltose and glucose has not been reported for other homologous enzymes. The crystal structure of TK-PUL has been determined at a resolution of 2.8 Å and represents the first analysis of a type III pullulan hydrolyse. The structure reveals that the last part of the N-terminal domain and the C-terminal domain are significantly different from homologous structures. In addition, the loop regions at the active-site end of the central catalytic domain are quite different. The enzyme has a well defined calcium-binding site and possesses a rare vicinal disulfide bridge. The thermostability of TK-PUL and its homologues may be attributable to several factors, including the increased content of salt bridges, helical segments, Pro, Arg and Tyr residues and the decreased content of serine.

## 1. Introduction

Pullulan, also known as $\alpha$-1,4-glucan or $\alpha$-1,6-glucan, is a polysaccharide synthesized from starch by the fungus *Aureobasidium pullulans* (Kim *et al.*, 1990). It is composed of repeating units of maltotriose linked by $\alpha$-1,6-glycosidic bonds or repeating units of isopanose joined by $\alpha$-1,4-glycosidic bonds (Leathers, 2003). The ratio of $\alpha$-1,4- to $\alpha$-1,6-glycosidic bonds in pullulan is 2:1. It is mainly used by cells to resist predation or desiccation, and is involved in the diffusion of molecules both into and out of cells. Pullulan has been used as a model substrate for studying starch-debranching enzymes (Plant *et al.*, 1986) and has applications in the food-processing and pharmaceutical industries (Shingel, 2004; Singh *et al.*, 2008).

Starches are a mixture of two high-molecular-weight polymers known as amylose and amylopectin (Swinkels, 1985). Amylose is a linear molecule formed by 100–10 000 D-glucose units connected by $\alpha$-1,4 linkages. In contrast, amylopectin, which constitutes approximately 73–80% of starch, consists of

24–30 $\alpha$-1,4-linked D-glucose units joined by $\alpha$-1,6-bonds, resulting in molecules with 9600–15 900 glucose units (Takeda *et al.*, 2003). Amylose and amylopectin possess a latent aldehyde group at the end of the polymeric chain that is known as the reducing end. Depending on their origin, starches have various industrial applications such as the manufacture of glucose and maltose syrups and the production of other oligosaccharides (van der Maarel *et al.*, 2002; Rendleman, 1997).

Owing to the complexity of its structure, depolymerization of starch into oligosaccharides and smaller sugars requires a range of enzymes including endoamylases, such as $\alpha$-amylases (EC 3.2.1.1), which cleave the chain internally, and exoamylases, such as glucoamylases (EC 3.2.1.3), which remove the terminal monosaccharides sequentially (Hii *et al.*, 2012). Transferases hydrolyse $\alpha$-1,4-glycosidic bonds and transfer part of the donor to form a new $\alpha$-1,4 link (*e.g.* amylomaltase; EC 2.4.1.25) or $\alpha$-1,6-glycosidic bond (*e.g.* branching enzyme; EC 2.4.1.18) with the acceptor. The converse reaction, namely the hydrolysis of $\alpha$-1,6-glycosidic bonds, is catalysed by the debranching enzymes and these are classified into the indirect and direct groups (Fogarty & Kelly, 1990). The indirect enzyme, amylo-1,6-glucosidase, requires the prior modification of the substrate by a transferase to leave a single $\alpha$-1,6-linked glucose moiety at the branch point. The direct enzymes, known as isoamylases and pullulanases, can hydrolyse $\alpha$-1,6-glycosidic bonds directly from unmodified substrate (Hii *et al.*, 2012). Hydrolysis of starch by $\alpha$-amylase produces limit dextrins and this greatly reduces glucose yield, although it can be improved significantly by the addition of pullulanases. In addition, by using pullulanase together with $\beta$-amylase in the starch-saccharification process, maltose yields can be increased by 20–25% (Poliakoff & Licence, 2007).

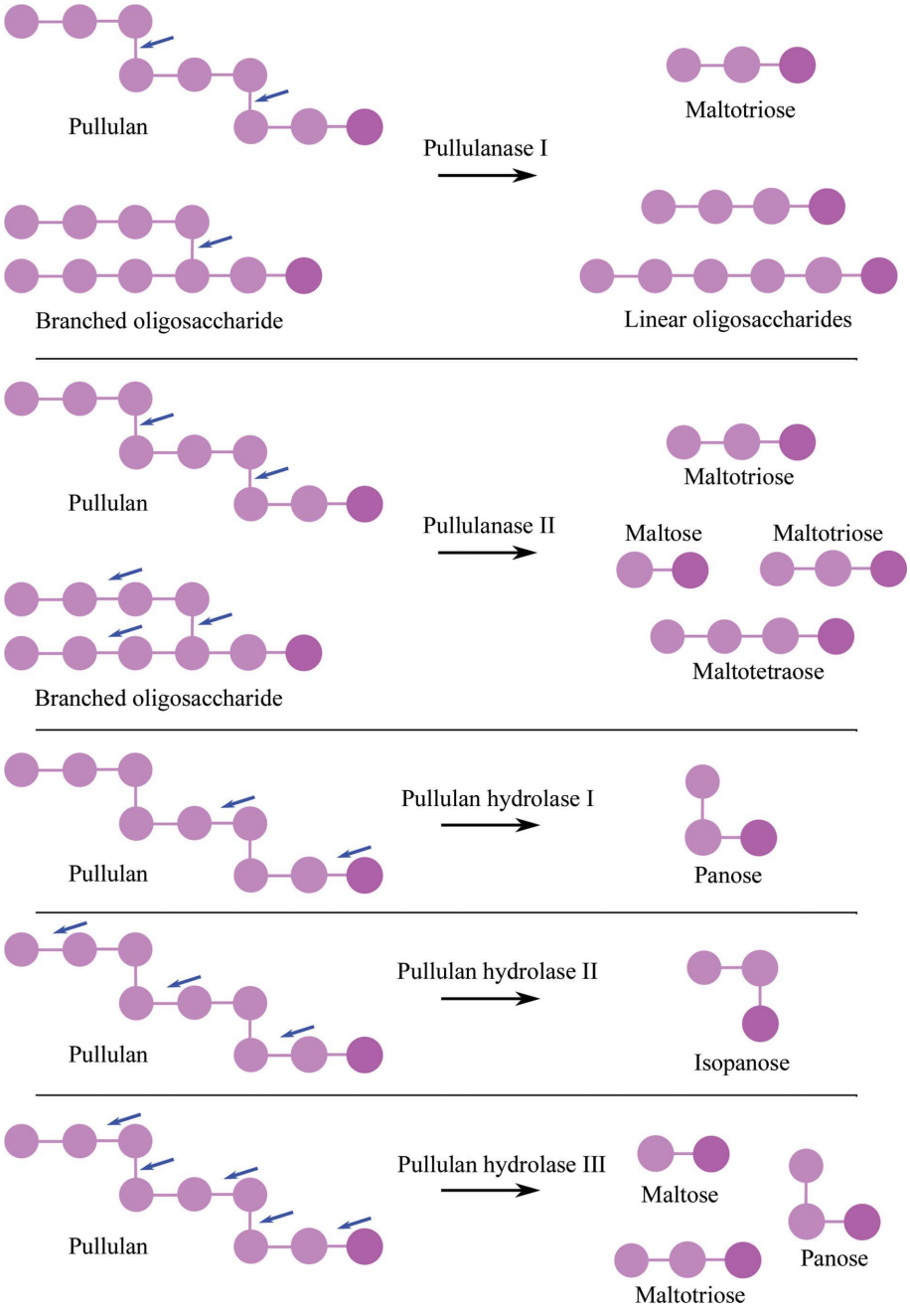


**Figure 1**
The reactions catalysed by different pullulan-hydrolyzing enzymes. The cleavage sites of pullulanases I and II as well as pullulan hydrolases I, II and III are indicated by arrows and the corresponding products are shown. The horizontal bars represent $\alpha$-1,4-glycosidic links and the vertical bars represent $\alpha$-1,6 links. The reducing ends of the oligosaccharides are indicated in darker purple. This figure is adapted from Nisha & Satyanarayana (2013).

Pullulanases, or more precisely pullulan-hydrolysing enzymes, are grouped into glycosyl hydrolase families 13 (GH13), 49 (GH49) and 57 (GH57) (Janeček *et al.*, 2014; MacGregor *et al.*, 2001) and, based on their substrate specificities and reaction products, are classified into five groups: pullulanases I and II, and pullulan hydrolases I, II and III (Hii *et al.*, 2012). Pullulanase I (EC 3.2.1.41), which was previously called R-enzyme, hydrolyses $\alpha$-1,6-glycosidic bonds in starch, pullulan, glycogen and limit dextrins, but does not degrade $\alpha$-1,4-glycosidic bonds. Pullulanase II or amylopullulanase (EC 3.2.1.1/41) acts on both $\alpha$-1,4 and $\alpha$-1,6 linkages in polysaccharides such as starch and limit dextrins, while it generally hydrolyses at $\alpha$-1,6-glycosidic bonds in pullulan, producing maltotriose (Nisha & Satyanarayana, 2013). Type I pullulan hydrolase (neopullulanase; EC 3.2.1.135) hydrolyses $\alpha$-1,4 bonds of pullulan to

generate panose and also hydrolyses both $\alpha$-1,4 and $\alpha$-1,6 linkages of starch and related polysaccharides with a low efficiency (Kuriki *et al.*, 1988). Pullulan hydrolase II (isopullulanase EC 3.2.1.57) cleaves the $\alpha$-1,4-linkages of pullulan and panose; however, it does not act on starch or dextran (Aoki & Sakano, 1997). Type III pullulan hydrolase cleaves both $\alpha$-1,4- and $\alpha$-1,6-glycosidic bonds of pullulan, resulting in the formation of maltotriose, panose, maltose and glucose. To date, only two enzymes in the latter class have been reported: those from *Thermococcus aggregans* (TA-PUL; Niehaus *et al.*, 2000) and *T. kodakarensis* (TK-PUL; Ahmad *et al.*, 2014). The characteristics of these pullulan-hydrolysing enzymes are summarized in Fig. 1.

The catalytic mechanism of pullulan-hydrolysing enzymes is similar to that of the $\alpha$-amylase family. The mechanism is characterized by $\alpha$-retaining double replacement involving two catalytic residues in the active site: a Glu as the acid–base catalyst and an Asp as the nucleophile (Fig. 2). Using TK-PUL residue numbering, the catalytic process involves Glu534 donating a proton to the glycosidic bond O atom and Asp601 nucleophilically attacking the C1 atom of glucose G1. This leads to the formation of an oxocarbonium ion-like transition state, followed by a covalent intermediate in which the G1 glycosyl moiety is covalently attached to Asp601. The covalent bond linking the G1 sugar to Asp601 is then cleaved following nucleophilic attack by a water or a glucose molecule, which ultimately replaces the protonated glucose G2 (Koshland, 1953; van der Maarel *et al.*, 2002). Asp503 is not directly involved in the catalytic process; instead, it binds to two hydroxy groups of the substrate and plays an important role in the distortion of the substrate (Uitdehaag *et al.*, 1999).

TK-PUL possesses both pullulanase and $\alpha$-amylase activities. The enzyme has a molecular mass of 84.4 kDa and consists of 748 amino acids. TK-PUL has highest activity at 95–100°C and optimal pH values of 3.5 and 4.2 in acetate and citrate buffers, respectively, although it is active in a broad pH range from 3.0 to 8.5. It does not require any metal ions for activity and shows a broad range of substrate specificity, including the ability to act on pullulan, $\beta$- and $\gamma$-cyclodextrin, starch, amylose, amylopectin, dextrin and glycogen (Ahmad *et al.*, 2014). Interestingly, cyclodextrins are well known competitive inhibitors of pullulanases (Duffner *et al.*, 2000) and none of the other enzymes have been reported to hydrolyse pullulan as efficiently. In addition, TK-PUL has the unique ability to hydrolyse maltotriose into maltose and glucose, which has not been reported for other homologous enzymes.

Pullulan-hydrolysing enzymes have high market value in the starch-saccharification industry for the production of glucose, maltose, maltotriose and maltotetraose syrups. They increase glucose and maltose production by 2% and 20–25%, respectively, and reduce the total reaction time and cost (Jensen & Norman, 1984). These enzymes are also employed as antistaling agents in the bread industry. Staling is owing to the retrogradation of amylopectin in starch and can be retarded by shortening the amylopectin chain length (Champenois *et al.*, 1999). These enzymes hydrolyse the branched maltodextrins produced by $\alpha$-amylase (Carroll *et al.*, 1987), which is required for making bread, and thereby help to eliminate 'gumminess'. They are also involved in the preparation of resistant starch and panose- and isopanose-containing syrups (Zhang & Jin, 2011; Machida *et al.*, 1986).

We have determined the crystal structure of TK-PUL at a resolution of 2.8 Å, with $R$ and $R_{\rm free}$ values of 24.2% and 27.9%, respectively. This represents the first structure of a type III pullulan hydrolyse and highlights a number of significant differences from other classes of pullulan-hydrolysing enzyme.

## 2. Methods

### 2.1. Crystallization

TK-PUL was expressed and purified using the methods described previously by Ahmad *et al.* (2014). In essence, the gene was amplified from genomic *T. kodakarensis* DNA and cloned into the pET-21a vector for expression in *Escherichia coli* BL21 CodonPlus (DE3)-RIL cells. Purification was conducted in 50 m*M* Tris–HCl pH 8.0 buffer and involved heat treatment at 80°C for 30 min to precipitate impurities, followed by ammonium sulfate precipitation and use of a Resource Q column (GE Healthcare, Piscataway, New Jersey, USA), eluting the protein with a 0–1 *M* NaCl gradient. Crystal

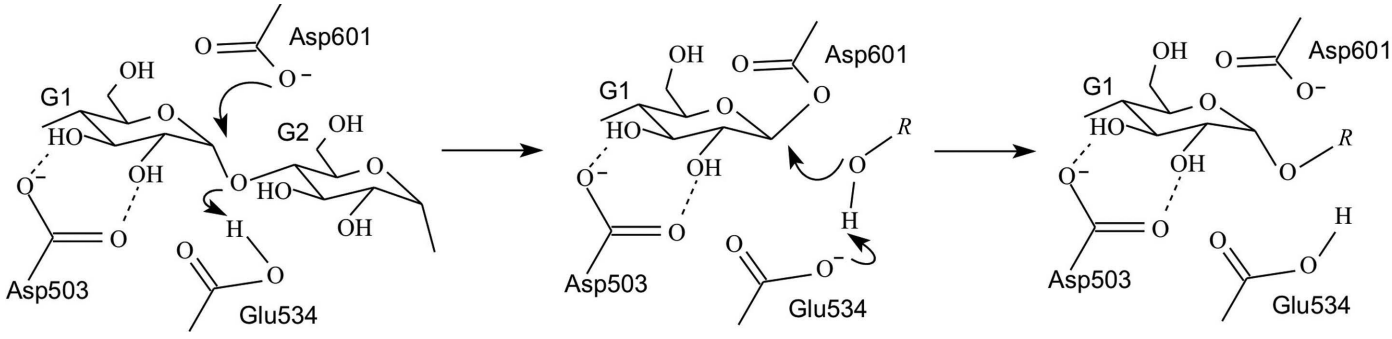


**Figure 2**
The catalytic mechanism of pullulan-hydrolysing enzymes. The amino acids are numbered according to the type III pullulan hydrolase from *T. kodakarensis* (TK-PUL). This figure is adapted from van der Maarel *et al.* (2002).

**Table 1**
X-ray statistics for the TK-PUL structure.

Values in parentheses are for the outer resolution shell.

| | |
|---|---|
| Beamline | I03, DLS |
| Wavelength (Å) | 0.9762 |
| Space group | *C*2 |
| Unit-cell parameters | |
| *a* (Å) | 192.6 |
| *b* (Å) | 63.9 |
| *c* (Å) | 56.1 |
| *β* (°) | 93.8 |
| Resolution range (Å) | 96.11–2.80 (2.95–2.80) |
| $R_{merge}$ (%) | 16.7 (90.0) |
| $R_{meas}$ (%) | 20.1 (106.2) |
| $CC_{1/2}$ (%) | 98.5 (53.6) |
| Completeness (%) | 99.8 (98.9) |
| Average $I/\sigma(I)$ | 5.4 (1.2) |
| Multiplicity | 3.3 (3.3) |
| No. of observed reflections | 56033 (7968) |
| No. of unique reflections | 17039 (2444) |
| Wilson plot *B* factor (Å$^2$) | 47.8 |
| Solvent content (%) | 52.8 |
| *R* factor (%) | 24.2 |
| $R_{free}$ (%) | 27.9 |
| R.m.s.d., bond lengths (Å) | 0.004 |
| R.m.s.d.,bond angles (°) | 0.621 |
| No. of reflections in working set | 16960 |
| No. of reflections in test set | 909 |
| Mean protein *B* factor (Å$^2$) | 47.3 |

screening was carried out using the sitting-drop method. Protein samples at 5, 10 and 16 mg ml$^{-1}$ were screened at 21°C with and without 1 m*M* $Ca^{2+}$, D-glucose, maltose, maltotriose, panose, *n*-dodecyl *α*-D-maltoside, *α*-cyclodextrin, *β*-cyclodextrin and *γ*-cyclodextrin (all in tenfold molar excess except for $Ca^{2+}$). Many crystal clusters were obtained in conditions D1, E1, F1, G1 and H1 from the MORPHEUS screen (Gorrec, 2009) and many single crystals for the enzyme in the presence of *n*-dodecyl *α*-D-maltoside were obtained in MORPHEUS condition B9. Further optimization revealed that better crystals could be obtained with the enzyme present at 10 mg ml$^{-1}$ in 0.02 *M* of each of sodium formate, ammonium acetate, trisodium citrate, sodium potassium tartrate (racemic) and sodium oxamate along with 0.1 *M* MES/imidazole pH 6.2, 14% PEG 500 MME and 7% PEG 20 000. Selected crystals were mounted in loops before flash-cooling in liquid nitrogen.

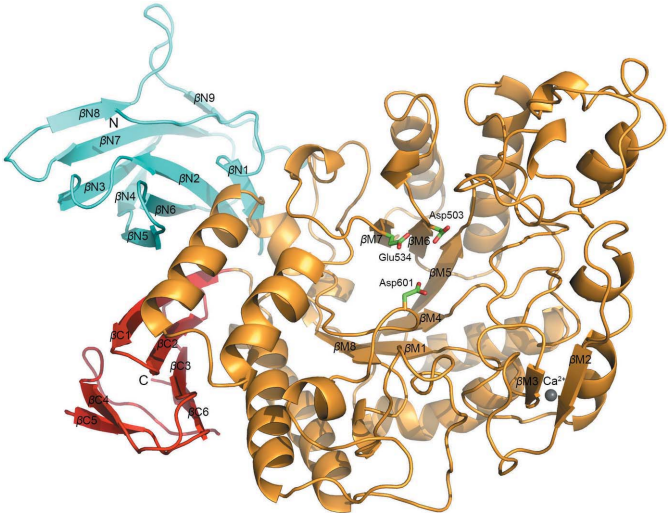


**Figure 3**
The crystal structure of TK-PUL. The N-terminal, central and C-terminal domains are coloured cyan, orange and red, respectively. The three residues forming the catalytic triad are shown in ball-and-stick representation and the calcium ion is shown as a grey sphere. The calcium-binding subdomain is visible on the lower right-hand side of the central domain and the kernel-like subdomain appears at the top right.

### 2.2. Data collection and data processing

X-ray diffraction data were collected at station I03 at Diamond Light Source (DLS) and the raw data were processed using *DIALS* (Winter *et al.*, 2018; Gildea *et al.*, 2014) to separate multiple lattices in space group *C*2. The space group was later confirmed by *POINTLESS* (Evans, 2006, 2011) and molecular replacement. Scaling and data reduction were carried out using *AIMLESS* (Evans & Murshudov, 2013) and data quality was checked by *phenix.xtriage* (Zwart *et al.*, 2005). *MATTHEWS_COEF* (Kantardjieff & Rupp, 2003, Matthews, 1968) suggested a solvent content of 52.8% with one molecule per asymmetric unit.

### 2.3. Structure determination

A partial structure solution for TK-PUL was identified by molecular replacement and refinement using the *MrBUMP* website (Keegan & Winn, 2008). The structure of *Thermus thermophilus* HB8 pullulanase (Ttha1563; PDB entry 2z1k; 37% sequence identity to TK-PUL; RIKEN Structural Genomics/Proteomics Initiative, unpublished work) was used as the search model and residues 286–764 were modelled into the electron density, which gave a *Phaser* log-likelihood gain (LLG) value of 235.0, a translation-function *Z*-score (TFZ) of 12.7 and an $R_{free}$ value of 44.7%. Since there was a large positive density at the N-terminal end of the partial structure, molecular replacement was repeated using *MOLREP* (Vagin & Teplyakov, 2010) with the partial solution as a fixed model and residues 103–220 of *Staphylothermus marinus* maltogenic amylase (SMMA; PDB entry 4aee; Jung *et al.*, 2012), which have approximately 10% sequence identity, as the search model for the missing region. This enabled the placement of an extra 101 residues (185–285) into the electron density, which decreased the $R_{free}$ value to 39.5%. The successful placement of this region required the use of the spherically averaged phased translation function, followed by a local phased rotation function and a phased translation function, as described by Vagin & Isupov (2001). It was not possible to build the first 184 N-terminal residues owing to a lack of electron density, and SDS–PAGE analysis gave a band of only 66 kDa, which suggested that these residues might have been cleaved during sample preparation or storage. Successive rounds of manual rebuilding and correction were carried out using *Coot* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010) interspersed with restrained refinement with *REFMAC* (Murshudov *et al.*, 2011) and *phenix.refine* (Adams *et al.*, 2010; Afonine *et al.*, 2012; Echols *et al.*, 2012). All statistics for data

collection, data processing, structure determination and refinement of the structure are shown in Table 1. The *VADAR* (Willard *et al.*, 2003) and *ESBRI* (Costantini *et al.*, 2008) online services were used to analyse hydrogen bonds, salt bridges and other factors related to thermostability of the enzymes. Figures were prepared using *PyMOL* (Schrödinger) and *CueMol* (http://www.cuemol.org). The structure and reflection files have been deposited in the PDB with accession code 5ot1.

## 3. Results and discussion

### 3.1. Tertiary structure of TK-PUL

Fig. 3 shows the tertiary structure of TK-PUL, which consists of an N-terminal domain, a central catalytic domain and a C-terminal domain. The N-terminal domain contains residues 185–280, which form an antiparallel $\beta$-barrel structure. The central catalytic domain adopts a TIM-barrel structure which is composed of residues 281–694. The C-terminal antiparallel $\beta$-barrel domain is formed by residues 694 onwards. The triangular assembly of the three domains places them spatially adjacent to each other.

Following the first $\beta$-strand in the central domain there is a region (300–350) inserted into the TIM barrel which contains two $\alpha$-helices, a $\beta$-hairpin and a few loops that are involved in binding a calcium ion. The third $\beta$-strand is followed by another insertion of 50 residues forming a very compact, kernel-like subdomain consisting of helical and $\beta$-strand elements. There are a few helical segments following the fifth and the sixth strands, and the last $\beta$-strand

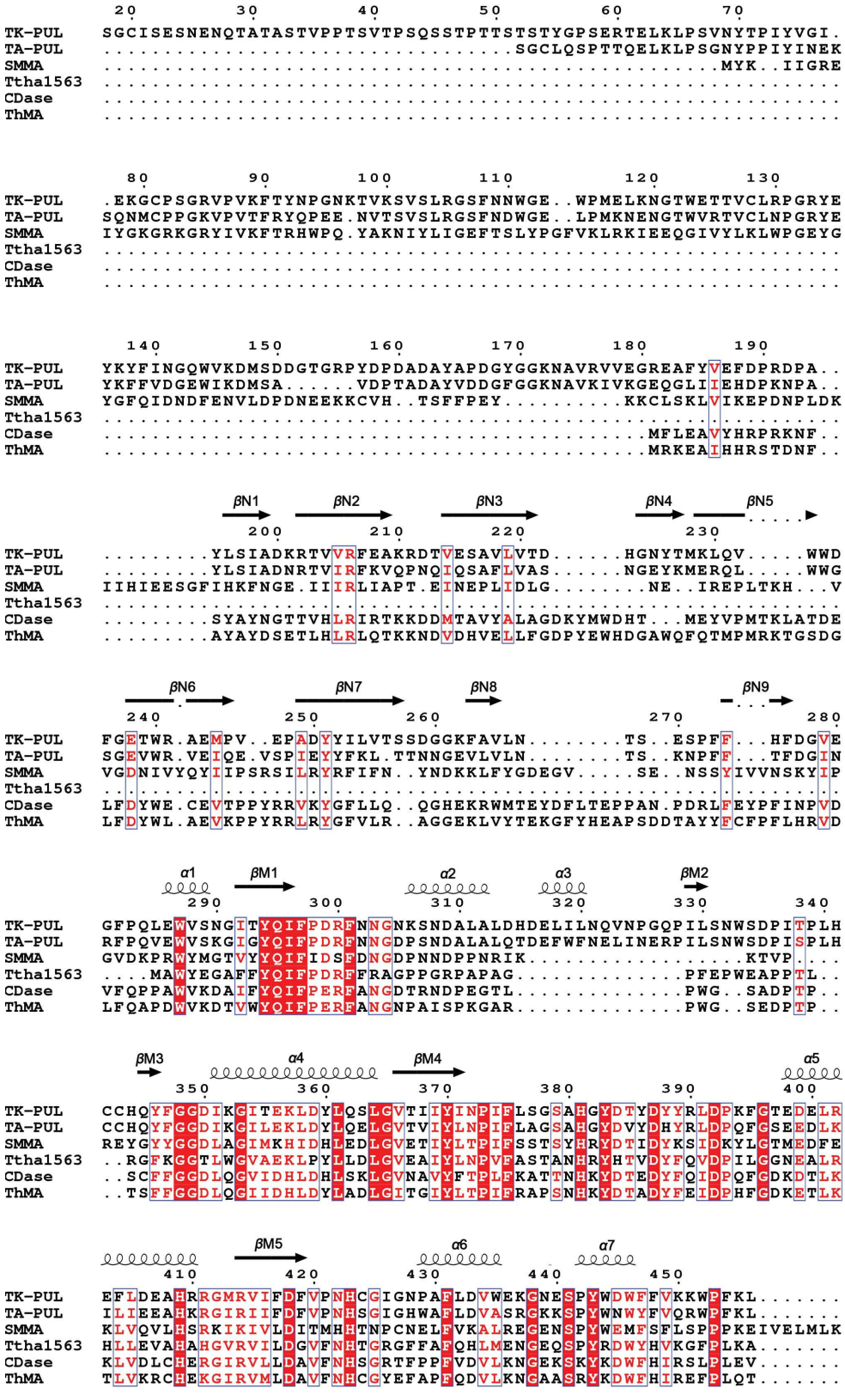


**Figure 4**
Sequence alignment and secondary structure of TK-PUL. The $\alpha$-helices and $\beta$-strands of TK-PUL are indicated, apart from for the N-terminal 184 residues which are missing in the structure. All conserved residues are boxed and the fully conserved residues are coloured white on a red background, while the conserved residues are coloured red. The sequences shown are for TK-PUL, the type III pullulan hydrolase from *Thermosphaera aggregans* (TA-PUL), maltogenic amylase from *Staphylothermus marinus* (SMMA), *Thermus thermophilus* HB8 pullulanase (Ttha1563), cyclomaltodextrinase from *Bacillus* sp. 1-5 (CDase) and maltogenic amylase from *Thermus* sp. IM6501 (ThMA). Alignment was performed using *ESPript* 3.0 (Gouet *et al.*, 2003; Robert & Gouet, 2014).

(the eighth) is followed by a flap-like structure which partially covers the active site and interacts with the kernel-like subdomain.

### 3.2. Structural comparison with homologues

A sequence alignment of TK-PUL with several related structures is shown in Fig. 4. The first 17 amino acids prior to the N-terminus of TK-PUL form a signal peptide which was included in the numbering scheme of Ahmad *et al.* (2014), as is adopted here. TK-PUL has 65.9% sequence identity to TA-PUL over 704 residues, whilst it shares only 27–37% sequence identity with a number of other homologues over approximately 480 residues which cover the central and C-terminal domains of TK-PUL. The N-terminal residues 185–280 of TK-PUL do not share high sequence similarity with the corresponding residues in SMMA; however, their tertiary structures are quite similar. Structure prediction using *HHpred* (Söding *et al.*, 2005) suggested that the preceding residues 78–180 are folded in a similar way to the corresponding region in the AMP-activated protein kinase from *Rattus norvegicus* (PDB entry 4yef; Mobbs *et al.*, 2015), although this region could not be modelled owing to the lack of electron density. This domain is also a member of the CBM48 (carbohydrate-binding module 48) family, the members of which include pullulanase, maltooligosyl trehalose synthase, starch branching enzyme, glycogen branching enzyme, glycogen debranching enzyme and isoamylase. In other microbial amylases, this region of the protein is referred to as the N′-domain and is involved in dimerization and the formation of a cap over the active-site cleft (Jung *et al.*, 2012). However, gel-filtration experiments demonstrated that recombinant TK-PUL exists in a monomeric form.

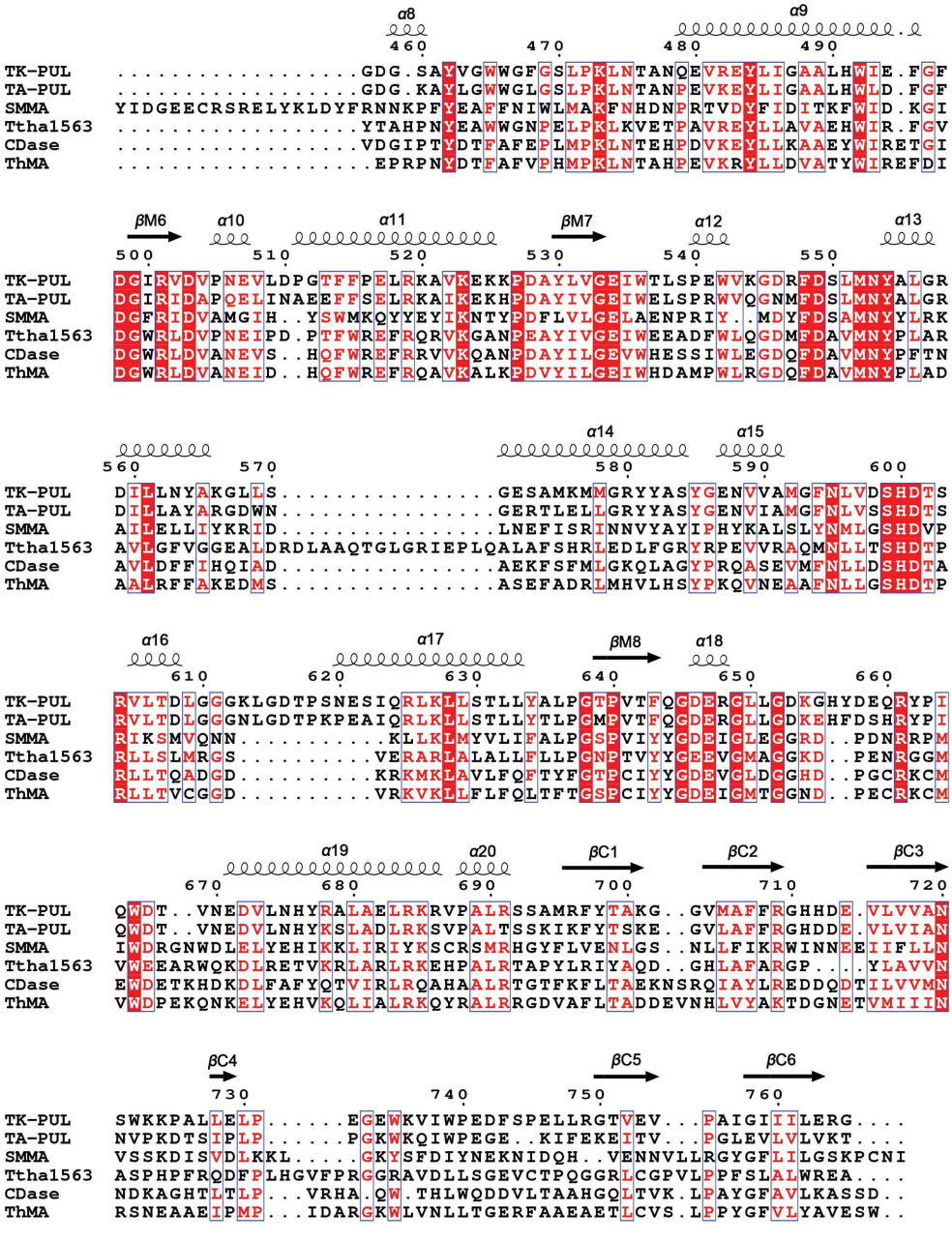


**Figure 4 (continued)**

Fig. 5 illustrates a structural superimposition of TK-PUL with the maltogenic amylases from *Thermus* sp. (ThMA; PDB entry 1sma; Kim *et al.*, 1999) and *Bacillus* sp. I-6 (PDB entry 1ea9; Lee *et al.*, 2002). The structures of both of these enzymes superpose with TK-PUL with r.m.s. $C^{\alpha}$ deviations of 2.1 Å, suggesting that the structures differ rather significantly from one another. However, superposition of just the large catalytic domains reduces the r.m.s. deviation to around 1.3 Å, suggesting that the other two, more peripheral domains differ somewhat in orientation from one structure to the next and give rise to the high overall r.m.s. deviation. Indeed, relative to PDB entry 1sma, the N-terminal domain has rotated by 20.5° and the C-terminal domain by 6.6°. Similar domain shifts were observed in a structural comparison with PDB entry 1ea9, where the N-terminal domain differs in orientation by 20.5° and the C-terminal domain by 9.6°. These domain movements were determined following superposition with the central catalytic domain and their magnitude may account for some of the

difficulties in the molecular-replacement analysis.

The TIM-barrel domain of TK-PUL shares high structural similarity with those of homologous proteins, while the N-terminal and C-terminal domains differ to a slightly greater extent. TK-PUL is also homologous to the type I pullulan hydrolases from *T. thermophilus* (PDB entry 2z1k; RIKEN Structural Genomics/Proteomics Initiative, unpublished work) and *B. stearothermophilus* (PDB entry 1j0h; Hondoh *et al.*, 2003) in both sequence and structure, ignoring that the former lacks the N-terminal domain. However, the loop regions at the active-site end of the catalytic domain are quite different in these structures. For example, the TK-PUL enzyme lacks the large insertion in the $\alpha$-helical region following the sixth strand which forms a flap over the active-site cavity in the *T. thermophilus* structure. The spatially adjacent loop following the seventh strand is substantially larger in the TK-PUL enzyme. The calcium-binding loop following the first $\beta$-strand of the TIM barrel is considerably larger in TK-PUL. This loop is oriented towards the active site and is likely to be involved in peripheral interactions with large oligosaccharide substrates.

### 3.3. Active site and specificity

The catalytic triad is composed of Asp503, Glu534 and Asp601 and is located at the base of a cavity that contains many exposed aromatic residues, as shown in Fig. 6. The very exposed aromatic side chains of Trp465 and Phe468 suggest a role in binding the hydrophobic faces of substrate sugars and have also been identified in related enzymes (Hondoh *et al.*, 2003; Ohtaki *et al.*, 2006). Critical involvement of aromatic residues in the catalytic activity of TK-PUL was also indicated during biochemical characterization upon the observation that the enzyme activity was strongly inhibited in the presence of *N*-bromosuccinimide, even at a final concentration of only 0.01% (Ahmad *et al.*, 2014).

Superposition of the structure with that of *B. subtilis* strain 168 type I pullulanase with maltose bound (PDB entry 2e9b; D. Malle, H. Iwamoto, Y. Katsuya, S. Utsumi & B. Mikami, unpublished work) indicated that two sugar rings could be accommodated well within the active site of TK-PUL. A maltose molecule in this position is within hydrogen-bonding distance of the catalytic residues and forms many hydrophobic interactions with aromatic side chains, including Trp465 and Phe468.

Of the related enzymes mentioned thus far, TK-PUL is unique in its ability to cleave $\alpha$-1,6-glycosidic bonds as well as $\alpha$-1,4 linkages in its linear substrate. The strong structural similarity between TK-PUL and it homologues in the vicinity of the catalytic centre means that this intriguing difference in specificity is difficult to account for in structural terms and is likely to stem from a range of subtle effects. However, TK-PUL and the related type I pullulan hydrolase from *T. aggregans* (TA-PUL in Fig. 4) both have a very substantial insertion of approximately 20 residues relative to the other homologues in the vicinity of residues 310–330 of TK-PUL. This insertion forms a large loop which projects towards the active-site region and is likely to affect peripheral interactions between the enzyme and its pullulan substrate. This loop, which can be clearly seen on the right-hand side of Fig. 5 (shown in green), contains a region of $\alpha$-helix at residues 310–320 followed by a proline-rich extended region at residues 325–330. The electron density for this loop suggests that it is very well defined and could well have a strong bearing on the specificity of the enzyme, possibly with respect to its ability to cleave $\alpha$-1,6-glycosidic linkages.

### 3.4. Calcium-binding loop and vicinal disulfide

The small subdomain formed by residues 300–350 is involved in calcium binding by octahedral coordination, as

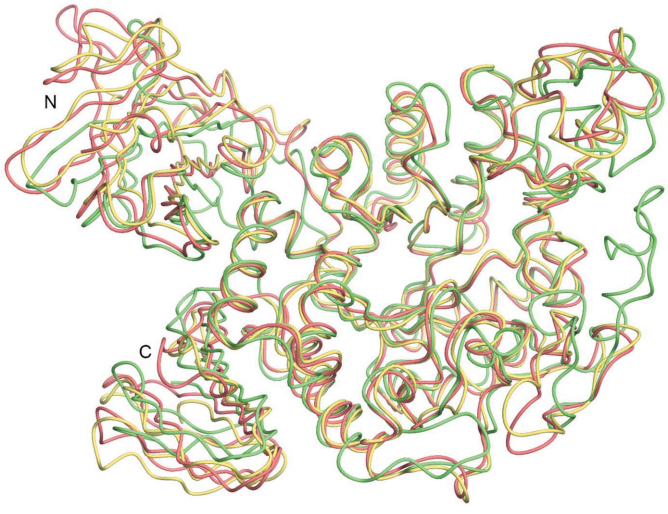


**Figure 5**
Superimposition of TK-PUL with two other homologues. TK-PUL, *Bacillus* sp. I-6 cyclomaltodextrinase (PDB entry 1ea9) and ThMA (PDB entry 1sma) are coloured green, yellow and pink, respectively.

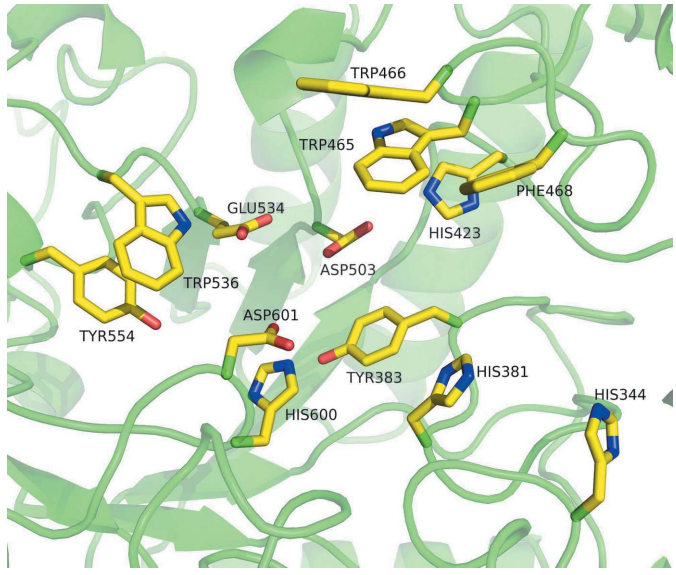


**Figure 6**
The active site of TK-PUL. The catalytic triad and aromatic residues that are involved in carbohydrate binding are shown in ball-and-stick representation.

shown in Fig. 7. This calcium-binding site has been reported in many other homologous proteins such as the R-47 $\alpha$-amylase II from *Thermoactinomyces vulgaris* (PDB entry 1wzk; M. Mizuno, K. Ichikawa, T. Tonozuka, A. Ohtaki, Y. Shimura, S. Kamitori, A. Nishikawa & Y. Sakano, unpublished work) and neopullulanase from *B. stearothermophilus* (PDB entry 1j0h; Hondoh *et al.*, 2003). The residues that participate in calcium binding, through their main-chain or side-chain atoms, include Asn303, Gly305, Asn308, Asp309, Gly348 and Asp350. The calcium ion is buried, completely dehydrated and could potentially play an important role in the stabilization of the protein. However, in contrast to the other known members of the $\alpha$-amylase family, there was no effect on the enzyme activity and thermal stability after extensive dialysis of the purified enzyme against 10 m*M* EDTA (Ahmad *et al.*, 2014; Ahmed *et al.*, 2016).

There is also a rare vicinal disulfide bridge in this region which is formed by Cys342 and Cys343. This may have an influence on the substrate specificity of TK-PUL, possibly in a redox-dependent manner, because the covalently linked cysteine side chains are oriented towards the active site. The finding that the sulfydryl-modifying reagents *p*-chloromercuribenzoic acid and iodoacetic acid did not result in significant inhibition of the activity of TK-PUL (Ahmad *et al.*, 2014) is consistent with the oxidized state of these two cysteines, which should preclude reaction with these two compounds. Given that TK-PUL is likely to be a secreted enzyme, oxidation of these two adjacent cysteines may occur *in vivo*.

### 3.5. Thermostability

The great thermostability of thermophilic proteins can be attributed to several factors. These proteins tend to have greater hydrophobicity (Haney *et al.*, 1997), a larger number of hydrogen bonds (Vogt *et al.*, 1997; Vogt & Argos, 1997) and salt bridges (Yip *et al.*, 1995, 1998; Haney *et al.*, 1997; Kumar *et al.*, 2000), increased helical content, a low occurrence of thermolabile residues such as cysteine and serine (Russell *et al.*, 1997), a high occurrence of Arg, Tyr and Pro residues (Watanabe *et al.*, 1997; Bogin *et al.*, 1998; Haney *et al.*, 1997), amino-acid substitutions within and outside the secondary structures (Zuber, 1988; Haney *et al.*, 1997; Russell *et al.*, 1997), better packing, smaller and less numerous cavities, deletion or shortening of loops (Russell *et al.*, 1997), increased surface area buried upon oligomerization (Salminen *et al.*, 1996) and increased polar surface area (Haney *et al.*, 1997; Vogt *et al.*, 1997; Vogt & Argos, 1997). However, it should be noted that no single factor proposed to contribute to protein thermostability is 100% consistent in all thermophilic proteins. Kumar *et al.* (2000) observed that the most consistent trend is shown by side chain–side chain hydrogen bonds and salt bridges. They may rigidify a thermophilic protein in the room-temperature range and allow it to be flexible enough at high temperature in order to function (Jaenicke & Böhm, 1998).

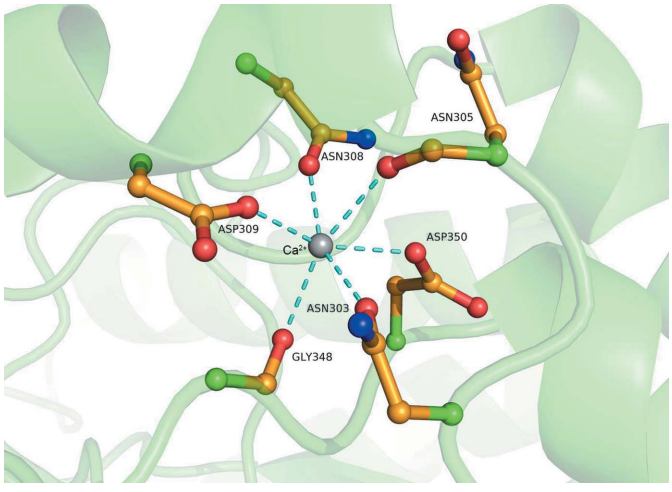


**Figure 7**
The calcium-binding site of TK-PUL. Several main-chain and side-chain carbonyl groups are involved.

**Table 2**
Thermostability-related factors for several thermophilic and mesophilic pullulan-hydrolyzing enzymes.

Factors which may contribute to the thermostability are shown in bold.

| | Thermophilic | | | | Mesophilic | | |
|---|---|---|---|---|---|---|---|
| Enzyme† | TK-PUL | 1sma | 1j0h | 2z1k | 2yoc | 2fh6 | 2wan |
| **Salt bridges (%)** | 25.7 | 29.9 | 30.6 | 30.1 | 20.4 | 17.6 | 16.5 |
| Hydrogen bonds (%) | 74 | 73 | 76 | 77 | 75 | 76 | 74 |
| **Helix content (%)** | 25 | 23 | 23 | 31 | 22 | 25 | 15 |
| **Proline content (%)** | 6.3 | 6.0 | 6.0 | 8.4 | 3.9 | 4.5 | 4.8 |
| **Arginine content (%)** | 4.8 | 6.0 | 5.8 | 8.4 | 3.9 | 4.1 | 2.1 |
| **Tyrosine content (%)** | 4.8 | 5.1 | 5.4 | 9.4 | 3.6 | 3.6 | 4.3 |
| Cysteine content (%) | 0.8 | 1.4 | 1.4 | 0.4 | 0.6 | 0.6 | 0.1 |
| **Serine content (%)** | 6.3 | 2.9 | 3.6 | 1.9 | 9.2 | 9.3 | 6.5 |

† All enzymes are represented by their PDB code (except for TK-PUL) as follows: 1sma, maltogenic amylase from *Thermus* sp.; 1j0h, type I pullulan hydrolase from *B. stearothermophilus*; 2z1k, type I pullulan hydrolase from *T. thermophilus*; 2yoc, pullulanase from *Klebsiella oxytoca* (East *et al.*, 2016); 2fh6, pullulanase from *K. aerogenes* (Mikami *et al.*, 2006); 2wan, pullulanase from *B. acidopullulyticus* (Turkenburg *et al.*, 2009).

A comparison of some of these factors for several thermophilic and mesophilic pullulan-hydrolyzing enzymes is shown in Table 2, and those which may contribute to the thermostability are shown in bold. Thermophilic enzymes show a higher content of salt bridges, helical segments and Pro, Arg and Tyr residues than the mesophilic enzymes, whilst the serine content tends to be much lower. However, the hydrogen-bond content is similar in all of the enzymes and the cysteine content is slightly higher in the thermophilic enzymes, both of which are not consistent with the trends mentioned above.

Kumar *et al.* (2000) suggested that the number of salt bridges in thermophilic and mesophilic homologues appears to correlate with the melting temperature $T_m$. In general, $\alpha$-helices enhance the rigidity and stability of proteins more than $\beta$-strands and loops. Helices in thermophilic enzymes favour arginine and avoid histidine and cysteine compared with those in mesophilic enzymes (Warren & Petsko, 1995; Kumar, Tsai *et al.*, 2000). The relatively high proline content of the thermophilic enzymes (Table 2) is consistent with the fact that this amino acid can only adopt a limited range of

conformations owing to the rigidity of the pyrrolidine ring. It was identified that thermophilic proteins tend to have proline as the second residue in $\beta$-turns or in the first turns of $\alpha$-helices (Bogin *et al.*, 1998; Watanabe *et al.*, 1997). Thus, proline is thought to exert crucial effects on protein thermostability by controlling the folding of the molecule. The high arginine and tyrosine content is expected to be significant in both short-range and long-range interactions owing to the large side chains of these amino acids. In addition, the guanidinium side chain of arginine can form salt bridges which stabilize proteins, and this effect is likely to be significant for TK-PUL. In contrast, the short side chains of Cys and Ser residues mostly only form local interactions and they are associated with deamidation reactions and oxidation at high temperatures (Russell *et al.*, 1997).

## 4. Summary

The crystal structure of a thermoacidophilic type III pullulan hydrolase, TK-PUL, has been determined at a resolution of 2.8 Å, representing the first structure of a type III pullulan hydrolase. The unique properties of TK-PUL, for example great thermostability and extraordinary stability over a broad pH range, make it an ideal enzyme for the starch industry. The first 184 residues are missing from the structure, which might be owing to cleavage during protein preparation. The structure of the last part of the N-terminal domain (185–280) and the C-terminal domain are slightly different from homologous structures, and the loop regions at the active-site end of the central catalytic domain are quite different, with a potential bearing on the unique specificity of this enzyme. The structure of the first part (residues 78–180) of the N-terminal domain is predicted to be similar to the corresponding region in the AMP-activated protein kinase from *R. norvegicus*. The region formed by residues 300–350 is involved in calcium binding, which has been reported in other homologous proteins. There is also a rare vicinal disulfide bridge formed by residues Cys342 and Cys343 which may have an influence on the substrate specificity. The thermostability of TK-PUL and its homologues may be attributed to several factors, including an increased content of salt bridges, helical segments, proline, arginine and tyrosine, and a decreased content of serine. Finally, the structure presented here contributes to the knowledge base of carbohydrate-metabolizing enzymes and sheds light on how their modular structures facilitate the recognition of and permit activity on different oligosaccharide motifs.

## Funding information

We gratefully acknowledge Diamond Light Source for X-ray beam time and travel support for data collection (award MX12342).

## References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Ahmad, N., Rashid, N., Haider, M. S., Akram, M. & Akhtar, M. (2014). *Appl. Environ. Microbiol.* **80**, 1108–1115.
Ahmed, N., Rashid, N., Haider, M. S. & Akhtar, M. (2016). US Patent US20140227744/US9340778.
Aoki, H. & Sakano, Y. (1997). *Biochem. J.* **323**, 859–861.
Bogin, O., Peretz, M., Hacham, Y., Korkhin, Y., Frolow, F., Kalb (Gilboa), A. J. & Burstein, Y. (1998). *Protein Sci.* **7**, 1156–1163.
Carroll, J. O., Boyce, C. O., Wong, T. M. & Starace, C. A. (1987). US Patent US4654216.
Champenois, Y., Della Valle, G., Planchot, V., Buleon, A. & Colonna, P. (1999). *Sci. Aliments*, **19**, 471–486.
Costantini, S., Colonna, G. & Facchiano, A. M. (2008). *Bioinformation*, **3**, 137–138.
Duffner, F., Bertoldo, C., Andersen, J. T., Wagner, K. & Antranikian, G. (2000). *J. Bacteriol.* **182**, 6331–6338.
East, A., Mechaly, A. E., Huysmans, G. H. M., Bernarde, C., Tello-Manigne, D., Nadeau, N., Pugsley, A. P., Buschiazzo, A., Alzari, P. M., Bond, P. J. & Francetic, O. (2016). *Structure*, **24**, 92–104.
Echols, N., Grosse-Kunstleve, R. W., Afonine, P. V., Bunkóczi, G., Chen, V. B., Headd, J. J., McCoy, A. J., Moriarty, N. W., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Adams, P. D. (2012). *J. Appl. Cryst.* **45**, 581–586.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Fogarty, W. M. & Kelly, C. T. (1990). *Microbial Enzymes and Biotechnology*, 2nd ed., pp. 71–132. Dordrecht: Springer.
Gildea, R. J., Waterman, D. G., Parkhurst, J. M., Axford, D., Sutton, G., Stuart, D. I., Sauter, N. K., Evans, G. & Winter, G. (2014). *Acta Cryst.* D**70**, 2652–2666.
Gorrec, F. (2009). *J. Appl. Cryst.* **42**, 1035–1042.
Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.
Haney, P., Konisky, J., Koretke, K., Luthey-Schulten, Z. & Wolynes, P. (1997). *Proteins*, **28**, 117–130.
Hii, S. L., Tan, J. S., Ling, T. C. & Ariff, A. B. (2012). *Enzym. Res.* **2012**, 921362.
Hondoh, H., Kuriki, T. & Matsuura, Y. (2003). *J. Mol. Biol.* **326**, 177–188.
Jaenicke, R. & Böhm, G. (1998). *Curr. Opin. Struct. Biol.* **8**, 738–748.
Janeček, Š., Svensson, B. & MacGregor, E. A. (2014). *Cell. Mol. Life Sci.* **71**, 1149–1170.
Jensen, B. & Norman, B. (1984). *Process Biochem.* **19**, 129–141
Jung, T.-Y., Li, D., Park, J.-T., Yoon, S.-M., Tran, P. L., Oh, B.-H., Janeček, Š., Park, S. G., Woo, E.-J. & Park, K.-H. (2012). *J. Biol. Chem.* **287**, 7979–7989.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 119–124.
Kim, J.-S., Cha, S.-S., Kim, H.-J., Kim, T.-J., Ha, N.-C., Oh, S.-T., Cho, H.-S., Cho, M.-J., Kim, M.-J., Lee, H.-S., Kim, J.-W., Choi, K.-Y., Park, K.-H. & Oh, B.-H. (1999). *J. Biol. Chem.* **274**, 26279–26286.
Kim, C.-H., Kim, D.-S., Taniguchi, H. & Maruyama, Y. (1990). *J. Chromatogr. A*, **512**, 131–137.
Koshland, D. E. (1953). *Biol. Rev.* **28**, 416–436.
Kumar, S., Ma, B., Tsai, C.-J. & Nussinov, R. (2000). *Proteins*, **38**, 368–383.
Kumar, S., Tsai, C.-J. & Nussinov, R. (2000). *Protein Eng.* **13**, 179–191.
Kuriki, T., Okada, S. & Imanaka, T. (1988). *J. Bacteriol.* **170**, 1554–1559.
Leathers, T. D. (2003). *Appl. Microbiol. Biotechnol.* **62**, 468–473.
Lee, H.-S., Kim, M.-S., Cho, H.-S., Kim, J.-I., Kim, T.-J., Choi, J.-H., Park, C., Lee, H.-S., Oh, B.-H. & Park, K.-H. (2002). *J. Biol. Chem.* **277**, 21891–21897.

Maarel, M. J. E. C. van der, van der Veen, B., Uitdehaag, J. C., Leemhuis, H. & Dijkhuizen, L. (2002). *J. Biotechnol.* **94**, 137–155.
MacGregor, E. A., Janeček, Š. & Svensson, B. (2001). *Biochim. Biophys. Acta*, **1546**, 1–20.
Machida, Y., Fukui, F. & Komoto, T. (1986). European Patent EP 0242459.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mikami, B., Iwamoto, H., Malle, D., Yoon, H.-J., Demirkan-Sarikaya, E., Mezaki, Y. & Katsuya, Y. (2006). *J. Mol. Biol.* **359**, 690–707.
Mobbs, J. I., Koay, A., Di Paolo, A., Bieri, M., Petrie, E. J., Gorman, M. A., Doughty, L., Parker, M. W., Stapleton, D. I., Griffin, M. D. W. & Gooley, P. R. (2015). *Biochem. J.* **468**, 245–257.
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.* D**67**, 355–367.
Niehaus, F., Peters, A., Groudieva, T. & Antranikian, G. (2000). *FEMS Microbiol. Lett.* **190**, 223–229.
Nisha, M. & Satyanarayana, T. (2013). *Bioengineered*, **4**, 388–400.
Ohtaki, A., Mizuno, M., Yoshida, H., Tonozuka, T., Sakano, Y. & Kamitori, S. (2006). *Carbohydr. Res.* **341**, 1041–1046.
Plant, A. R., Morgan, H. W. & Daniel, R. M. (1986). *Enzyme Microb. Technol.* **8**, 668–672.
Poliakoff, M. & Licence, P. (2007). *Nature (London)*, **450**, 810–812.
Rendleman, J. A. (1997). *Biotechnol. Appl. Biochem.* **26**, 51–61.
Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J. & Taylor, G. L. (1997). *Biochemistry*, **36**, 9983–9994.
Salminen, T., Teplyakov, A., Kankare, J., Cooperman, B. S., Lahti, R. & Goldman, A. (1996). *Protein Sci.* **5**, 1014–1025.
Shingel, K. I. (2004). *Carbohydr. Res.* **339**, 447–460.
Singh, R. S., Saini, G. K. & Kennedy, J. F. (2008). *Carbohydr. Polym.* **73**, 515–531.
Söding, J., Biegert, A. & Lupas, A. N. (2005). *Nucleic Acids Res.* **33**, W244–W248.
Swinkels, J. J. M. (1985). *Starch*, **37**, 1–5.
Takeda, Y., Shibahara, S. & Hanashiro, I. (2003). *Carbohydr. Res.* **338**, 471–475.
Turkenburg, J. P., Brzozowski, A. M., Svendsen, A., Borchert, T. V., Davies, G. J. & Wilson, K. S. (2009). *Proteins*, **76**, 516–519.
Uitdehaag, J. C., Mosi, R., Kalk, K. H., van der Veen, B. A., Dijkhuizen, L., Withers, S. G. & Dijkstra, B. W. (1999). *Nature Struct. Mol. Biol.* **6**, 432–436.
Vagin, A. A. & Isupov, M. N. (2001). *Acta Cryst.* D**57**, 1451–1456.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Vogt, G. & Argos, P. (1997). *Fold. Des.* **2**, S40–S46.
Vogt, G., Woell, S. & Argos, P. (1997). *J. Mol. Biol.* **269**, 631–643.
Warren, G. L. & Petsko, G. A. (1995). *Protein Eng. Des. Sel.* **8**, 905–913.
Watanabe, K., Hata, Y., Kizaki, H., Katsube, Y. & Suzuki, Y. (1997). *J. Mol. Biol.* **269**, 142–153.
Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D. & Wishart, D. S. (2003). *Nucleic Acids Res.* **31**, 3316–3319.
Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Cryst.* D**74**, 85–97.
Yip, K. S. P., Britton, K. L., Stillman, T. J., Lebbink, J., de Vos, W. M., Robb, F. T., Vetriani, C., Maeder, D. & Rice, D. W. (1998). *FEBS J.* **255**, 336–346.
Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. & Rice, D. W. (1995). *Structure*, **3**, 1147–1158.
Zhang, H. & Jin, Z. (2011). *Carbohydr. Polym.* **83**, 865–867.
Zuber, H. (1988). *Biophys. Chem.* **29**, 171–179.
Zwart, P., Grosse-Kunstleve, R. & Adams, P. (2005). *CCP4 Newsl. Protein Crystallogr.* **43**, contribution 7.