Title	A novel glycoside hydrolase family 97 enzyme: Bifunctional -l-arabinopyranosidase/ -galactosidase from <i>Bacteroides thetaiotaomicron</i>
Author(s)	Kikuchi, Asako; Okuyama, Masayuki; Kato, Koji; Osaki, Shohei; Ma, Min; Kumagai, Yuya; Matsunaga, Kana; Klahan, Patcharapa; Tagami, Takayoshi; Yao, Min; Kimura, Atsuo
Citation	Biochimie, 142, 41-50 https://doi.org/10.1016/j.biochi.2017.08.003
Issue Date	2017-11
Doc URL	http://hdl.handle.net/2115/72070
Rights	© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/
Rights(URL)	http://creativecommons.org/licenses/by-nc-nd/4.0/
Type	article (author version)
File Information	kikuchi_et_al Supplementary_materials (2).pdf

[Instructions for use](#)

A novel glycoside hydrolase family 97 enzyme: bifunctional β -L-arabinopyranosidase/ α -galactosidase from *Bacteroides thetaiotaomicron*

Asako Kikuchi^{a,†}, Masayuki Okuyama^{a,*,†}, Koji Kato^b, Shohei Osaki^a, Min Ma^a, Yuya Kumagai^a, Kana Matsunaga^a, Patcharapa Klahan^a, Takayoshi Tagami^a, Min Yao^b, Atsuo Kimura^{a,*}

^a Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan.

Kita-9 Nishi-9, Kita-ku, Sapporo 060-8589, Japan.

^b Faculty of Advance Life Science, Hokkaido University, Sapporo, Japan.

Kita-10 Nishi-8, Kita-ku, Sapporo 060-0810, Japan.

*Corresponding authors: M Okuyama: Address, Kita-9 Nishi-9, Kita-ku, Sapporo 060-8589, Japan; E-mail, okuyama@abs.agr.hokudai.ac.jp; Tel, +81-11-706-2816.

A Kimura: Address, Kita-9 Nishi-9, Kita-ku, Sapporo 060-8589, Japan; E-mail, kimura@abs.agr.hokudai.ac.jp; Tel/Fax, +81-11-706-2808.

[†] These authors contributed equally to this work.

ABSTRACT

Glycoside hydrolase family 97 (GH97) is one of the most interesting glycosidase families, which contains inverting and retaining glycosidases. Currently, only two enzyme types, α -glucoside hydrolase and α -galactosidase, are registered in the carbohydrate active enzyme database as GH97 function-known proteins. To explore new specificities, BT3661 and BT3664, which have distinct amino acid sequences when compared with that of GH97 α -glucoside hydrolase and α -galactosidase, were characterized in this study. BT3664 was identified to be an α -galactosidase, whereas BT3661 exhibits hydrolytic activity toward both β -L-arabinopyranoside and α -D-galactopyranoside, and thus we designate BT3661 as a β -L-arabinopyranosidase/ α -D-galactosidase. Since this is the first dual substrate specificity enzyme in GH97, we investigated the substrate recognition mechanism of BT3661 by determining its three-dimensional structure and based on this structural data generated a number of mutants to probe the enzymatic mechanism. Structural comparison shows that the active-site pocket of BT3661 is similar to GH97 α -galactosidase BT1871, but the environment around the hydroxymethyl group of the galactopyranoside is different. While BT1871 bears Glu361 to stabilize the hydroxy group of C6 through a hydrogen bond with its carboxy group, BT3661 has Asn338 at the equivalent position. Amino acid mutation analysis indicates that the length of the side chain at Asn338 is important for defining specificity of BT3661. The $k_{\text{cat}}/K_{\text{m}}$ value for the hydrolysis of *p*-nitrophenyl α -galactoside decreases when Asn338 is substituted with Glu, whereas an increase is observed when the mutation is Ala. Interestingly, mutation of Asn338 to Ala reduces the $k_{\text{cat}}/K_{\text{m}}$ value for hydrolysis of *p*-nitrophenyl β -L-arabinopyranoside.

Keywords

Glycoside hydrolase family 97, β -L-arabinopyranosidase/ α -galactosidase, *Bacteroides thetaiotaomicron*, crystal structure, substrate recognition

46 **Abbreviations**

47 HPAEC-PAD, high-performance anion-exchange chromatography with pulsed amperometric
48 detection; PNP, *p*-nitrophenol; PNP-Ara, PNP β -L-arabinopyranoside; PNP-Gal, PNP α -D-
49 galactopyranoside; PNP-Glc, PNP α -D-glucopyranoside; PNP-Xyl, PNP α -D-xylopyranoside.

50

1. Introduction

A large number of glycoside hydrolases have been conscientiously classified into 145 glycoside hydrolase families (GHs), with some of these families deleted from the carbohydrate-active enzyme (CAZy) database (<http://www.cazy.org>) [1]. Since CAZy classification is based on amino acid sequence similarity, the database enables an understanding of evolutionary relationships in the same GH enzymes, and facilitates prediction of structural features and enzymatic functions of uncharacterized members. A certain GH contains obvious homologs with high sequence identity and also proteins with low sequence similarity. In general, conservation of function fluctuates in a cluster with 20–30% identity [2–4] and, in the case of GHs, a changes in the primary structure is often associated with the divergence of substrate specificity [5].

Although sequence similarity is low, crucial amino acids required for catalysis and the catalytic mechanism are generally preserved within a GH. However, GH97 contains inverting and retaining glycosidases and is one of the unique GHs. As another example, GH23 is known to include inverting lysozymes and retaining lytic transglycosylases. GH97, comprising five subfamilies (97a–97e), was established through sequence analyses [6]. Divergence of the catalytic mechanisms in GH97 has been elucidated through structural and biochemical analysis [7,8]. The mechanisms are governed by the arrangement of the catalytic residues on the “ β -face” of the substrate. HHET and HE motifs are located at the C-terminus of β -strands 3 and 5 of the $(\beta/\alpha)_8$ barrel catalytic domain of the inverting enzymes [7], and Glu residues in the motifs form the catalytic general base of an inverting mechanism [7,8]. Retaining enzymes have a conserved Asp residue that provides a catalytic nucleophile for the retaining mechanism at the C-terminus of β -strand 4 [9]. An additional characteristic of GH97 enzymes is the coordination of a calcium ion at the active center. The calcium ion facilitates the departure of a leaving group and stabilizes the deprotonation state of the base catalyst in an inverting enzyme [10].

Most of GH97 enzymes are distributed in the Bacteroidetes phylum. According to the Pfam database (<http://pfam.xfam.org/>), 718 out of 1035 sequences in bacteria are derived from this

phylum. *Bacteroides thetaiotaomicron*, which is known to be involved in the breakdown of plant polysaccharides and host-derived glycans in the colon [11–13], possesses 10 GH97 proteins. Among them, one inverting α -glucoside hydrolase [7,8] and a small number of retaining α -galactosidases [9,14] have been characterized. The inverting enzyme, categorized into subfamily 97a, is a member of the starch utilization system (*sus*) and catalyzes the hydrolysis of α -glucosidic linkages of maltooligosaccharides, which are produced by the action of SusA and SusG [8,15]. This enzyme has several appellations, including 97A1_BACTH, SusB, *BtGH97a*, and BT3703, which can cause confusion. We use the most commonly used locus-tag name in this paper. The two retaining α -galactosidases, BT1871 and BT2620, which are members of subfamily 97b, are believed to be involved in the degradation of the α -galactosidic linkage in galactomannan. Substrate specificity of BT1871 has been examined for its transglycosylation properties [16]. BT1871 uses *p*-nitrophenyl α - and β -mannosides as acceptor molecules and has a subsite structure to accommodate α -galactosyl and α/β -mannosyl moieties. BT2620 forms part of a polysaccharide utilization locus (PUL; MAN-PUL1) and has been proposed to degrade a terminal α -galactosidic linkage in yeast α -mannan [14]. Apart from Bacteroidetes, a halophilic GH97 inverting α -glucoside hydrolase from the Arctic marine bacterium (*Pseudoalteromonas* sp. K8) has been found to prefer the α -1,6-glucosidic linkage to the α -1,4-glucosidic linkage, which is distinct from BT3703 [17]. The tolerance to high concentrations of salts is due to a chloride-ion binding site where a unique Arg residue plays a central role [18]. Only these enzymes have well characterized structures and functions, with most GH97 proteins uncharacterized.

Novel specificities are frequently discovered within existing GHs. Among the evolutionarily related GHs to GH97, namely GH27, 31 and 36 [6–8,16], GH31 is such a family that consists of enzymes with distinct functions. The family comprises not only hydrolases and transglycosylases but also an α -1,4-glucan lyase [19,20]. In addition, several enzymes with unique specificity, including 1,4- α -glucan 4- α -glucosyltransferase [21], α -galactosidase [22], sulfoquinovosidase [23], cycloalternan-specific α -1,3-glucosidase [24,25] and dextranase [26], have been found in

GH31 over the past years. These findings in the related family suggest the possibility that GH97 also contains an enzyme exhibiting unpredictable specificity.

This study aimed to identify a GH97 enzyme with novel specificity. We have interest in two *B. thetaiotaomicron* GH97 paralogs, BT3661 and BT3664, which are categorized into subfamily 97c. These are annotated as putative α -glucosidases in UniProtKB. This study revealed that BT3661 and BT3664 are a bifunctional β -L-arabinopyranosidase/ α -galactosidase and an α -galactosidase, respectively. In particular, BT3661 was demonstrated as the first specificity in GH97. We also determined the crystal structure of BT3661, which enabled us to interpret its novel specificity in GH97.

2. Materials and methods

2.1. Sequence analysis

The amino acid sequences of the catalytic domain of GH97 enzymes, collected from the seed alignment in Pfam PF10566 (<http://pfam.xfam.org/family/PF10566>), and the GH97 paralogs in *B. thetaiotaomicron* were aligned by Clustal omega [27] at the MPI Bioinformatics Toolkit server (<https://toolkit.tuebingen.mpg.de/#/>) [28]. A maximum likelihood phylogeny was estimated from the alignment by PhyML-SMS [29]. The bootstrap consensus tree inferred from 100 replicates was taken as the evolutionary history. Amino acid sequences with Asp, corresponding to a catalytic nucleophile residue, were manually collected from the members of the seed alignment in Pfam PF10566 as representative sequences of GH97 retaining glycosidases. Multiple alignment of full-length sequences and three-dimensional structures were made by MAFFTash (<https://sysimm.ifrec.osaka-u.ac.jp/MAFFTash/>) (Supplementary Fig. 2). A dendrogram of the GH97 retaining enzymes was calculated using PhyML, as described above. Graphical representation of the sequence conservation was made using Skylign [30]. The presence and location of secretion signal peptide cleavage sites in the BT3661 and BT3664 sequences were predicted by Phobius [31]. BT3661 and BT3664 subcellular localizations were predicted by

Loctree3 [32] at the Predict Protein server (<https://predictprotein.org/home>).

2.2. Gene cloning

Each gene encoding BT3661 or BT3664 was amplified from chromosomal DNA prepared from *B. thetaiotaomicron* ATCC 29148 by PCR using the primers for BT3661 [5'-ATGAAGAACTTGTAAGCACT-3' (sense) and 5'-CTACTTTATACAAAGAATACCTCC-3', (antisense)] or for BT3664 [5'-ATGAAACATTCCTTTAATTGTTTA-3' (sense) and 5'-TTACTTTATTACTGCCACAAATCC-3' (antisense)]. The purified PCR products were cloned into the pBluescript II SK (+) vector (Stratagene, Agilent Technology, Santa Clara, CA, USA), and designated as SK-*bt3661* and SK-*bt3664*, respectively. For construction of pCold-*bt3661*, the gene coding for residues 20–647 of BT3661 was amplified by PCR from SK-*bt3661* using primers with restriction cleavage sites: 5'-CCGGGCGAGCTCCAGGATGTGGTTGTAAAAGGA-3' (sense; *SacI*-site underlined) and 5'-GCCCCGGCTCGAGCTACTTTATACAAAGAATACC-3' (antisense; *XhoI*-site underlined). The PCR product was restricted with *SacI* and *XhoI*, and cloned into appropriately digested pCold I DNA (Takara Bio, Shiga, Japan). Sequence analysis revealed a single base difference compared with the GenBank accession number NC_004663, which resulted in an F478L mutation. For construction of pCold-*bt3664*, the sequence coding for BT3664 residues 29–638 was amplified by PCR from SK-*bt3664* using primers: 5'-GCGGAGCTCGATACAGGTGTGATTTCT-3' (sense; *SacI*-site underlined) and 5'-GCGAAGCTTTTATTACTTTATTACTGCCAC-3' (antisense; *HindIII*-site underlined). The resultant PCR product was restricted with *SacI* and *HindIII*, and cloned into appropriately digested pColdI plasmid DNA. PrimeSTAR Max DNA polymerase (Takara Bio) was used for all PCR experiments according to the manufacturer's instruction.

2.3. Gene expression and protein purification

E. coli BL21(DE3) cells harboring pCold-*bt3661* or pCold-*bt3664* were cultured in LB

medium ($10 \text{ g}\cdot\text{L}^{-1}$ Bacto tryptone, $5 \text{ g}\cdot\text{L}^{-1}$ Bacto yeast extract and $5 \text{ g}\cdot\text{L}^{-1}$ NaCl) supplemented with $0.05 \text{ mg}\cdot\text{mL}^{-1}$ ampicillin at 37°C to mid-exponential phase ($\text{OD}_{600} \approx 0.7$). The culture was cooled on ice for 30 min, at which time isopropyl β -D-thiogalactopyranoside was added to a final concentration of 0.2 mM and cells were incubated at 12°C for induction of BT3661 or at 15°C for BT3664. The induction temperature was optimal for each recombinant protein production. After 24 h with vigorous shaking, the cells were collected by centrifugation ($5,000 \times g$, 5 min, 4°C) and resuspended in 20 mM Hepes-NaOH, 300 mM NaOH, pH 7.5 (buffer-A), followed by disruption through sonication. The supernatant was obtained by centrifugation ($20,000 \times g$, 20 min, 4°C).

Recombinant proteins were purified from the above supernatant, loaded onto a Ni^{2+} -chelating Sepharose Fast Flow column (GE healthcare, Chicago, IL, USA), which was prepared according to the supplier's protocol and pre-equilibrated with buffer-A. After washing the column with buffer-A containing 50 mM imidazole, the absorbed proteins were eluted with a linear gradient of 50 to 300 mM imidazole in buffer-A. Active fractions were passed through a Sephadex G-25 (GE Healthcare) column pre-equilibrated with 50 mM sodium acetate, 100 mM NaCl and 25 mM CaCl_2 , pH 5.5 (buffer-B). The obtained fractions were pooled and concentrated using a 50 kDa MWCO Vivaspin 20 (Sartorius, Göttingen, Germany). For preparation of protein crystallization, the protein was further subjected to a Sephacryl S-200 (GE Healthcare) gel-filtration column pre-equilibrated with buffer-B. The active fractions were pooled, dialyzed against 20 mM sodium acetate buffer containing 10 mM CaCl_2 , pH 5.5, and concentrated using a 50 kDa MWCO Vivaspin 20. Fractions containing BT3661 were determined by specific activity and SDS-PAGE. All purification steps were performed at 4°C . The concentration of the purified protein was estimated by amino acid analysis of a protein hydrolysate (6 M HCl, 110°C , 24 h) using a JLC-500/V (JOEL, Tokyo, Japan) equipped with a ninhydrin-detection system. The purification procedures for BT1871 were described in our previous report [9].

2.4. Enzyme assay

The hydrolysis activity toward each *p*-nitrophenyl (PNP) glycoside, PNP β -L-arabinopyranoside (PNP-Ara) (Carbosynth Ltd., Berkshire, UK), PNP α -D-galactopyranoside (PNP-Gal) (Tokyo Chemical Industry Co., LTD., Tokyo, Japan), PNP α -D-glucopyranoside (PNP-Glc) (Nacalai Tesque, Kyoto, Japan) and PNP α -D-xylopyranoside (PNP-Xyl) (Santa Cruz Biotechnology, Inc. Co., Santa Cruz, CA, USA) was determined at 35 °C in 50 mM sodium acetate, pH 5.5, containing 10 mM CaCl₂ and 0.002 mg·mL⁻¹ bovine serum albumin (BSA), and various concentrations of substrates. The reactions were stopped by mixing with two volumes of 1 M Na₂CO₃/50 mM EDTA·2Na. The amount of *p*-nitrophenol released was measured by absorption at 400 nm in a 1-cm cuvette, using a molar extinction coefficient of 5560 M⁻¹·cm⁻¹.

The pH-dependent hydrolytic rate of BT3661 and its variants was measured at 2 mM PNP-Ara for wild-type (84 nM), N338E (35 nM) and N338E/Q378A (160 nM) or 2 mM PNP-Gal for N338A (46 nM), N338D (198 nM), Q378A (162 nM) and N338D/Q378A (112 nM) in the following buffers: 50 mM sodium acetate (pH 3.6–6.0), 50 mM Mes-NaOH (pH 5.0–7.0), 50 mM Hepes-NaOH (pH 6.5–9.0) and 50 mM glycine-NaOH (9.5–12.5), all of which contained 10 mM CaCl₂ and 0.002 mg·mL⁻¹ BSA. The susceptibility of BT3661 to pH was evaluated as follows: the enzyme (84 nM) was incubated in the above-mentioned buffers for 24 h at 4 °C and then the hydrolytic rate was measured using 2 mM PNP-Ara as the substrate. To examine the effect of calcium ions, the enzyme activity toward 2 mM PNP-Ara was measured with various concentrations of EDTA·2Na (0–3 mM). An appropriate volume of 0.5 M EDTA-NaOH solution (pH 6.0) was added to make the reaction mixture. The enzyme concentration used was 177 nM. The recovery of enzyme activity by addition of calcium ions was evaluated by measuring the activity of the enzyme (177 nM) under various concentrations of CaCl₂ (0–100 mM) in 50 mM sodium acetate buffer pH 5.5, containing 1 mM EDTA·2Na, 0.002 mg·mL⁻¹ BSA and 2 mM PNP-Ara. The kinetic constants were calculated by fitting the Michaelis-Menten equation by nonlinear regression using the computer program GraFit 7 (Erithacus Software Ltd., Surrey, UK). When k_{cat}

and K_m could not be estimated due to limited solubility of the substrate, the k_{cat}/K_m was evaluated at low substrate concentrations (1–2.5 mM).

2.5. Enzyme activity on natural polysaccharides

For the analysis of the hydrolysis of gum arabic from the acacia tree (Nacalai tesque) or arabinogalactan from larch wood (Tokyo Chemical Industry Co., LTD.), the polysaccharide (40 mg·mL⁻¹) was incubated with BT3661 (480 nM for gum arabic and 32 nM for arabinogalactan) at 35 °C in 40 mM sodium acetate (pH 5.5) containing 10 mM CaCl₂ and 0.002 mg·mL⁻¹ bovine serum albumin (BSA). The reaction was terminated by heating at 100 °C for 5 min. The reaction products were analyzed by thin-layer chromatography (TLC) and high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). TLC was performed using 0.25-mm layers of silica gel 60 F₂₅₄ (Merck, Darmstadt, Germany). The sugars were developed twice using a solvent system of acetonitrile:water (4:1, v/v) and visualized by spraying with an α -naphthol–sulfuric acid solution (15% sulfuric acid in methanol containing 0.3 mg/mL α -naphthol), followed by heating at 110 °C for 5 min. The concentrations of L-arabinose and D-galactose were measured by HPAEC–PAD using a Dionex ICS-3000 system (Dionex/Thermo Fisher Scientific, Idstein, Germany). The analytical column (CarboPac PA1 analytical column, 4 mm I.D. × 250 mm; Dionex/Thermo Fisher Scientific) was equilibrated with 48 mM NaOH at a flow rate of 0.8 mL·min⁻¹. Samples containing xylose or sorbitol as an internal standard were isocratically eluted with 48 mM NaOH. Sodium hydroxide solutions were prepared from super special grade 50% NaOH solution (Wako Pure Chemical Industries, Ltd., Osaka, Japan). The concentration of each saccharide was calculated from chromatographic peak areas using a calibration curve prepared from known saccharide concentrations and an internal standard. Chromatograms were evaluated with Chromeleon software (Dionex/Thermo Fisher Scientific). The activity of BT3664 (731 nM) toward polygalaturonic acid (Nacalai Tesque) or pectin (Tokyo Chemical Industry Co., LTD.) was analyzed by TLC using the same procedure as described above.

2.6. Site-directed mutagenesis

Point mutations to BT3661 were introduced by site-directed mutagenesis using the methods for the PrimeSTAR mutagenesis basal kit (Takara Bio). The oligonucleotides used for each mutation were as follows: for N338E, 5'-ATCGAT**GA**ATGGTGGGATACTAGAATC-3' (sense; substituted codon in boldface) and 5'-CCACCAT**TC**ATCGATCAACGCATACTC-3' (antisense); for N338A, 5'-ATCGAT**GC**CTGGTGGGATACTAGAATC-3' (sense) and 5'-CCACCAG**GC**ATCGATCAACGCATACTC-3' (antisense); for N338D, 5'-ATCGAT**GA**CTGGTGGGATACTAGAATC-3' (sense) and 5'-CCACCAG**TC**ATCGATCAACGCATACTC-3' (antisense); for Q378A, 5'-ATTGAAG**CG**GGACCGGTGAACCGAATG-3' (sense) and 5'-CGGTCCC**G**CTTCAATGTCATTCCAATA-3' (antisense). The mutant enzymes were prepared from *E. coli* transformants harboring each expression plasmid, followed by the aforementioned purification procedure.

2.7. Crystallization and data collection

Crystallization screening was performed using a series of crystallization kits from Qiagen (Hilden, Germany) by the sitting-drop vapor-diffusion method, in which 0.75 μL of 10 $\text{mg}\cdot\text{mL}^{-1}$ protein solution (20 mM sodium acetate containing 10 mM CaCl_2 , pH 5.5) was mixed with an equal volume of reservoir solution. The appropriate crystals of BT3661 for structural analysis were obtained within 2 weeks at 20 $^{\circ}\text{C}$ under a reservoir solution containing 0.2 M lithium sulfate, 0.1 M Mes-NaOH buffer (pH 6.0) and 45% (v/v) 2-methyl-2,4-pentanediol.

Because the crystallization conditions of BT3661 contained 45% (v/v) 2-methyl-2,4-pentanediol, which can be used as a cryoprotectant, the crystals were directly selected from crystallization drops and flash cooled for X-ray diffraction experiments. Diffraction data were collected on beamline BL-41XU at SPring-8 (Hyogo, Japan) under a stream of liquid nitrogen at

100 K. The dataset was indexed, integrated, scaled and merged using the *XDS* program suite [33]. The asymmetric unit of BT3661 contained four molecules corresponding to a Matthews coefficient [34] of $2.8 \text{ \AA}^3 \text{ Da}^{-1}$ and an estimated solvent content of 56.4%. The statistics of data collection and processing are summarized in Table 1.

2.8. Structure determination and refinement

The structure of Bt3661 was determined by the molecular replacement method with the program *AutoMR* in the *Phenix* program package [35,36] using the structure of GH97 α -galactosidase (PDB ID: 3A24) as a search model. Several rounds of refinement were performed using the program *Phenix.refine* in the *Phenix* program suite, alternating with manual fitting and rebuilding based on $2F_o - F_c$ and $F_o - F_c$ electron density using *COOT* [36,37]. The final refinement statistics and geometry defined by MolProbity [38] are shown in Table 1. PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.) was used to generate figures of the solved structure.

3. Results

3.1. Sequence analysis of GH97 paralogs in *B. thetaiotaomicron*

A phylogenetic tree of the catalytic domains of GH97 members was constructed to search for a sequence evolutionarily far from function-known proteins (Supplementary Fig. 1). Among the 10 paralogs in *B. thetaiotaomicron*, BT3661 and BT3664 classified into GH97 subfamily 97c [6] have distinct amino acid sequences compared with that of the function-known proteins, BT3073, BT1871 and BT2620. BT3661 and BT3664 have an Asp residue conserved in retaining GH97 enzymes [9] but do not have HHET and HE motifs found in inverting enzymes [7], which suggests that BT3661 and BT3664 are retaining glycosidases. The subcellular localization of BT3661 and BT3664 were predicted by LocTree3. Localization of both proteins was predicted to be secreted: the space external to the outermost structure of a cell. The secretion signal peptides of these two

enzymes were predicted by Phobius. BT3661 and BT3664 were predicted to have a secretion signal sequence composed of 19 and 27 residues at the N-terminus.

3.2. Production of BT3664

Recombinant BT3664 was successfully produced in *E. coli* transformed with pCold-*bt3664*. Substrate specificity of the recombinant enzyme purified by Ni²⁺-affinity chromatography was examined with 2 mM PNP-Gal, PNP-Ara, PNP-Glc and PNP-Xyl (Table 2). The enzyme exhibited substantial activity toward PNP-Gal, which is ~50 times larger than that toward PNP-Ara, and minimal activity towards PNP-Glc and PNP-Xyl. No product was detected on the TLC plate from the reaction mixtures containing BT3664 and each polysaccharide (polygalacturonic acid or pectin). These results indicate that BT3664 is an α -galactosidase.

3.3. Production and characterization of BT3661

Recombinant BT3661 was produced using the expression plasmid pCold-*bt3661*, which encodes mature BT3661 with an N-terminal (His)₆-tag. The hydrolytic activities toward 2 mM PNP glycosides were determined (Table 2). The enzyme hydrolyzed PNP-Ara and PNP-Gal with velocities of 0.558 and 0.208 $\mu\text{mol}/\text{min}/\text{mg}$, respectively, but no activity was observed towards PNP-Glc or PNP-Xyl. The optimum pH for the hydrolysis reaction was 5.5. After treatment at various pH values for 24 h at 4 °C, BT3661 maintained more than 90% of its hydrolytic activity when compared with that of the original enzyme over the pH range of 4–12. In the presence of 1 mM EDTA·2Na, the enzyme lost catalytic activity and this was restored by the addition of 5 mM CaCl₂. Kinetic parameters of BT3661 for hydrolysis of PNP-Ara and PNP-Gal were determined (Table 3). The $k_{\text{cat}}/K_{\text{m}}$ value of BT3661 for PNP-Ara was 2.8 times greater than that for PNP-Gal. The kinetic parameters of BT1871 toward PNP-Ara, which exhibited a $k_{\text{cat}}/K_{\text{m}}$ value of 778 s⁻¹ mM⁻¹ for PNP-Gal [9], were determined to know the hydrolytic activity of an unequivocal α -galactosidase toward β -L-arabinopyranoside. The constants were 15.9 ± 1.8 mM and 68.9 ± 5.1

s⁻¹ for K_m and k_{cat} , respectively, and the k_{cat}/K_m value for PNP-Ara (4.36 s⁻¹ mM⁻¹) was ~1/180 of that for PNP-Gal. BT3661 exhibited the distinct substrate specificity of BT1871 and we thus regarded BT3661 as a bifunctional β -L-arabinopyranosidase/ α -D-galactopyranosidase.

The hydrolytic reaction of BT3661 on polysaccharide substrates, gum arabic from the acacia tree and arabinogalactan from larch wood, was investigated. The release of arabinose and galactose from gum arabic and that of arabinose from arabinogalactan were observed by TLC (Fig. 1). Release rates of these monosaccharides were calculated through HPAEC-PAD analysis. The rates of arabinose and galactose release from gum arabic were 2.02 and 1.43 μ mol/min/mg ($v/[E]_0 = 2.47$ and 1.75 s⁻¹, respectively). The release rate of arabinose from arabinogalactan was 40.9 μ mol/min/mg ($v/[E]_0 = 50.0$ s⁻¹), but that of galactose was below the detection limit.

3.4. Crystal structure of BT3661

The crystal structure of BT3661 was determined at 2.3 Å resolution by the molecular replacement method using the structure of BT1871 (PDB ID: 3A24 [3]) as the search model. The structure was refined to crystallographic R_{work} and R_{free} values of 20.7% and 25.3%, respectively. The model showed well-defined geometric parameters based on the Ramachandran plot (favored region: 95.6%, allowed region: 4.2%, outliers region: 0.2%). The crystal contained four BT3661 molecules in the asymmetric unit (residues 16–647 for chain A, residues 20–647 for chain B, residues 16–647 for chain C and residues 19–647 for chain D were built, respectively). The structures of the four molecules were very similar with RMSDs (root mean square deviations) of 0.13–0.21 Å among 628–632 C α atoms of respective chains. As the electron density for chain A was more clearly visible than that for the other molecules, we have described the structure based on chain A. The BT3661 monomer structure contains three domains (Fig. 2), which is similar to other GH97 members: the (β/α)₈ barrel-fold catalytic domain (domain A, residues 301–554) is located between two β -sandwich folded domains, the N-terminal domain (domain N, residues 20–300) and the C-terminal domain (domain C, residues 555–647).

A structural comparison with a catalytic nucleophile mutant BT1871 (BT1871-D415G) in complex with β -lactosyl α -galactoside (PDB ID: 5E1Q [16]) was performed by TM-align (<http://zhanglab.ccmb.med.umich.edu/TM-align/>) [39] to understand the substrate-binding features of BT3661. The comparison showed that the overall structure of BT3661 was fairly similar to that of BT1871-D415G with a TM-score and an RMSD value of 0.86644 and 2.73 Å for 595 residues, respectively. The residues of the catalytic nucleophile and general acid/base are Asp408 and Glu463, respectively, and a calcium-ion, to which three glutamates coordinate, Glu179, Glu457 and Glu463, were almost identical to that of BT1871. The residues, Trp304, Trp306, Asp337, Trp366, Lys406, Phe409, His438, Glu457 and Glu463, which would be associated with the stabilization of a sugar moiety at the -1 subsite through a hydrogen bond or van der Waals contacts, were also placed with positions and orientations similar to BT1871 (Fig. 3A). In contrast, residues involved in the stabilization of a C6 hydroxy group of the α -galactosyl moiety were different between BT1871 and BT3661 (Fig. 3B). The former used Glu351 to stabilize this hydroxy group through a hydrogen bond with its carboxy group, whereas BT3661 used Asn338. Almost half of the representative sequences of GH97 retaining glycosidases, which are collected from the seed alignment in Pfam PF10566, possess a Glu residue at this equivalent position and the sequence with Asn represents a smaller part of the group (Fig. 3C). For the sequences recorded in UniRef50_F9D4D2, which is a cluster of amino acid sequences that have at least 50% sequence identity to and 80% overlap with BT3661, approximately one-third of the sequences possess an Asn equivalent to Asn338. Other sequences in this cluster have Ala, Gly, or Ser rather than Asn, but no sequence was found to have Glu at this position (Fig. 4). The side chain of Glu351 of BT1871 forms a hydrogen bond with His245 on a β -hairpin (His233–Ala260) of the domain N (Fig. 3B). BT3661 had no corresponding β -hairpin, but the $\beta \rightarrow \alpha$ loop 3 (Ser369–Met384) of the catalytic $(\beta/\alpha)_8$ barrel domain was located in the corresponding position and Gln378 was in close proximity to both Asn338 and a C6 hydroxy group of the sugar moiety at the -1 subsite. A conservation of the respective regions among GH97 retaining glycosidases, the β -

hairpin of BT1871 and the $\beta \rightarrow \alpha$ loop 3 of BT3661, is extremely low, and the pairs, Glu351–His245 (BT1871) and Asn338–Gln378 (BT3661), are not common in GH97 retaining glycosidases (Fig. 3C). All sequences that have Asn equivalent to Asn338 in UniRef50_F9D4D2 have Gln corresponding to Gln378 (Fig. 4).

Glu179 and Phe467 were placed near the +1 subsite, which is similar to BT1871 (Fig. 5A). In addition, Trp171 is involved in the formation of the +1 subsite and forms a clamp-like subsite structure together with Phe467. An aromatic residue equivalent to Trp171 was conserved in only sequences that had a shorter evolutionary distance to BT3661 (Fig. 5C). BT1871 has no corresponding tryptophan residue and no such aromatic clamp (Fig. 5B). In Fig. 3B and 5A, β -L-Arap (1 $\rightarrow$ 3) L-Araf, which should be a part of a natural substrate of BT3661, was virtually placed in the active-site pocket. The L-Araf moiety appears to be stabilized by the aromatic clamp. While the plus subsites of BT1871 are mainly composed of Trp474 of the $\beta \rightarrow \alpha$ loop 6 of the catalytic domain (Fig. 5B), those of BT3661 had a slit-like pocket that is narrow because part of a long loop (Pro163–Glu180) extends from the domain N and the $\beta \rightarrow \alpha$ loop 6 (Fig. 5A). A multiple sequence alignment indicates a low conservation of the relevant region (Pro163–Glu180) (Fig. 5C), suggesting the divergence of the structure of the plus subsites in GH97 retaining enzymes.

3.5. Substrate specificity of mutant BT3661s

Mutational studies were conducted to understand the specificity of BT3661. We selected Asn338 as a target residue for mutagenesis. As mentioned above, Asn338 is equivalent to Glu351 in BT1871 and is directly involved in the stabilization of the C6 hydroxy group of the substrate through a hydrogen bond. Three mutant enzymes, N338E, N338A and N338D, were generated to examine the role of Asn338. Glu, Ala and Asp are found at the corresponding position in GH97 retaining glycosidases (Fig. 3C). In addition, Gln378, which is in close structural proximity to the side chain of Asn338 and appears to be in the vicinity of the C5 hydroxymethylene (Fig. 3B), was also selected for mutagenesis. Thus, Q378A, N338E/Q378A and N338D/Q378A mutants were

also prepared. The latter two double-mutant enzymes were prepared to ensure that the carboxy group of Glu338 or Asp338 was in the same orientation as that of Glu351 of BT1871. The optimum pH values of these mutant enzymes were 5.5–5.7, and almost identical to the wild-type BT3661. Kinetic parameters of the mutant enzymes for hydrolysis of PNP-Ara and PNP-Gal were determined (Table 3). The parameters for several mutant enzymes were not determined because the K_m values were beyond the solubility limit of the substrates; however, the specific constants, k_{cat}/K_m , were estimated at low substrate concentrations. Substitution of Asn338 with Glu affected the hydrolytic activity toward PNP-Gal more significantly than to PNP-Ara. The K_m value of N338E for PNP-Gal was beyond the measurement limit and its k_{cat}/K_m exhibited 10% that of the wild-type enzyme. The K_m value for PNP-Ara was also greater than the measurable limit but its k_{cat}/K_m was almost equivalent to that of the wild-type enzyme, indicating an increase in both k_{cat} and K_m . The N338A mutation substantially reduced the hydrolytic activity toward β -L-arabinopyranoside, but an increase toward α -galactoside was observed, resulting in a decrease in the k_{cat}/K_m value by 3.6% and an increase by 860%, respectively. The mutation of Asn338 to Asp gave a slightly negative effect on β -L-arabinopyranosidase activity, but exhibited a positive effect on α -galactosidase activity. Replacement of Gln378 for Ala had negligible effect on the hydrolytic activity towards PNP-Ara, but a 4-fold rise in the k_{cat}/K_m value was observed for PNP-Gal. The double mutant enzymes, N338E/Q378A and N338D/Q378A, had very similar k_{cat}/K_m values towards the substrates as the two corresponding single mutations, N338E and N338D.

4. Discussion

To find an enzyme with novel activity in GH97, we have focused on two GH97 proteins, BT3661 and BT3664, which are categorized into a different subfamily, 97c, from function-known sequences [6]. BT3664 was identified to be an α -galactosidase in this study. A gene encoding BT3664 is located in RG-II PUL 3 (PUL64) together with genes encoding GH43_34 endo-1,4- β -xylanase, GH43_10 β -xylosidase, GH29 α -L-fucosidase and SusCD [6,40,41]. RG-II PUL 3 is

expressed in response to rhamnogalacturonan II [42] and BT3664 could be involved in the degradation of this plant polysaccharide. However, it is uncertain how BT3664 is associated with the degradation of this polysaccharide, because rhamnogalacturonan II consists of α -(1 $\rightarrow$ 4) linked galacturonic residues and several side chains with 10 different glycosyl residues, but has no α -galactosidic residue [43,44].

We demonstrate that BT3661 possesses β -L-arabinopyranosidase activity. Compared with unambiguously confirmed β -L-arabinopyranosidases from *Streptomyces avermitilis* [45] and *Geobacillus stearothermophilus* T6 [46], BT3661 displays promiscuous specificity. These two GH27 enzymes display negligible activity toward α -galactoside and much greater k_{cat}/K_m values for hydrolysis of PNP-Ara than that of PNP-Gal, whereas BT3661 exhibits only 2.8 times higher k_{cat}/K_m values towards PNP-Ara than that towards PNP-Gal. We thus consider BT3661 as a bifunctional β -L-arabinopyranosidase/ α -galactosidase. BT3661 catalyzes the release of arabinose/galactose and arabinose from gum arabic and arabinogalactan, respectively (Fig. 2). Although gum arabic from *Acacia* species is considered not to contain α -galactopyranoside [47–50], Aspinall and Knebl demonstrated that an α -galactopyranose residue occupies the terminal positions of gum arabic through controlled partial acid hydrolysis, α -galactosidase-treatment and a lectin-binding assay [51]. BT3661 is thus likely to be involved in the degradation of gum arabic and arabinogalactan. Moreover, the $v/[E]_0$ values for the polysaccharides were rather slow, which can possibly be explained by the number of β -L-arabinopyranosides and α -galactosides in polysaccharides being conceivable small; i.e., the substrate concentration may be too low. Alternatively, BT3661 is considered not to have specificity for polysaccharides. The enzyme might prefer oligosaccharides produced by the actions of endoglycanases. The genes for BT3661 and BT3664 are located close to each other, but the former is not included in the RG-II PUL 3 [40,41]. Thus, the physiological role of BT3661 is likely to be unrelated to that of BT3664. Conceivably, the gene for BT3661 was generated by a gene duplication of the gene for BT3664, and BT3661 has obtained specificity toward β -L-arabinopyranosides and α -galactosides through

distinct molecular evolution from BT3664.

Since BT3661 exhibits the dual substrate specificity, which is found in GH97 for the first time, and to understand its substrate recognition mechanism, its crystal structure was determined and compared with the structure of BT1871, which strictly recognizes α -galactosidase. While both enzymes have similar active-site pocket structures (Fig. 3A), different structural features were found around a hydroxymethyl group of α -galactopyranoside (Fig. 3B). BT1871 stabilizes a C6 hydroxy group of the α -galactoside through a hydrogen bond with a Glu residue (Glu351) [16], whereas BT3661 possesses Asn338 instead of this Glu residue. The mutational study suggests that the bifunctional substrate specificity of BT3661 relies on the length of the side chain of Asn338 and van der Waals interactions. Replacement of Asn338 by Glu, which was performed to mimic BT1871, substantially decreases the specificity toward α -galactoside; although, the negative charge at this location has a positive effect on its specificity, as shown by the N338D mutation (discussed below). The Ala mutation increases specificity to α -galactoside. These results indicate that a longer side chain at position Asn338 prevents binding of α -galactoside, whereas a shorter side chain creates additional space within the substrate pocket to accommodate the 6-CH₂OH of α -galactoside. Similarly, a substitution of Gln378 for Ala that increases the k_{cat}/K_m value 4-fold seems to afford space for the hydroxymethyl group at C5. N338A displayed a decrease in activity toward PNP-Ara, of which the k_{cat}/K_m value for hydrolysis was reduced to 3.6% that of the wild-type enzyme. A reduction in the size of the side chain may facilitate weaker binding of β -L-arabinopyranoside because of the loss of a van der Waals contact. Substitution of Asn338 for Asp showed 50% reduction in the k_{cat}/K_m value for hydrolysis of PNP-Ara. The size of the side chain is near identical to that of Asn, but the negative charge of the carboxy group may destabilize the interaction with L-arabinopyranoside. The substitution results yielded an increase in the k_{cat}/K_m value for hydrolysis of PNP-Gal. The carboxy group of the Asp residue possibly forms a hydrogen bond with the OH-6 and promotes charge separation ($O6^{\delta-} \cdots H6^{\delta+}$). Electron donation from the polarized OH-6 may stabilize an oxocarbenium-ion like transition state of the

hydrolysis of α -galactoside. Double-mutant enzymes, N338E/Q397A and N338D/Q397A, were constructed to orient the carboxy side chains to match that of Glu361 of BT1871. The k_{cat}/K_m values of these mutant enzymes are, however, similar to their single mutant counterparts, N338E and N338D, suggesting that the orientation of the carboxy groups remains unchanged and the effect of Asn338 on the specificity is larger than that of Gln379.

We show that BT3661 catalyzes the hydrolysis of both α -galactoside and β -L-arabinopyranoside. Such dual substrate specificity has not been shown previously for GH97 enzymes. We also identify that Asn338 is responsible for this bifunctional specificity by solving the crystal structure and using site-directed mutagenesis. In UniRef50_F9D4D2, approximately one-third of the sequences possess an Asn equivalent to Asn338 (Fig. 4) and these are expected to be bifunctional β -L-arabinopyranosidase/ α -galactosidases. Moreover, another one-third of the sequences have Ala instead of Asn (Fig. 4) and these are likely to be α -galactosidases based on our mutagenesis study (Table 3). As an example, BT3664 with Ala334 at the equivalent position displays substantial specificity toward α -galactoside (Table 2). UniRef50_F9D4D2 is a cluster that has at least 50% sequence identity with BT3661 and can be expected to have similar enzymatic function [2,3]; however, the cluster is found to contain at least two kinds of enzyme with different substrate specificity. Clearly the functions of most proteins for which amino acid sequences have been disclosed by advances in genome sequence analysis remain unknown. The present study shows that substantive analysis of proteins is often required for clarifying their functions.

Author contributions

M. Okuyama conceived and coordinated the study, analyzed data, and wrote the paper. A. Kikuchi, S. Osaki, and K. Matsunaga performed the kinetic experiments. K. Kato and T. Tagami performed the crystal structure analysis. M. Ma, Y. Kumagai, and P. Klahan provided technical assistance. M. Yao and A. Kimura provided essential equipment and facilities. All authors

reviewed the results and approved the the manuscript.

Conflict of interest

The authors declare that they have no conflicts of interest resulting from the contents of this article.

Acknowledgements

We thank the staff of the Instrumental Analysis Division of the Creative Research Institution at Hokkaido University for the amino acid analysis. We thank the staff of the beamline BL41XU (proposal No. 2017A2545) at SPring-8 for their assistance during data collection. We thank the Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

Funding

This work was supported by JSPS KAKENHI [grant number 26450114].

References

- [1] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, *Nucleic Acids Res.* 42 (2014) D490-5. doi:10.1093/nar/gkt1178.
- [2] A.E. Todd, C.A. Orengo, J.M. Thornton, Evolution of function in protein superfamilies, from a structural perspective, *Journal of Molecular Biology.* 307 (2001) 1113–1143. doi:10.1006/jmbi.2001.4513.
- [3] C.A. Wilson, J. Kreychman, M. Gerstein, Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores, *Journal of Molecular Biology.* 297 (2000) 233–249. doi:10.1006/jmbi.2000.3550.

- 519 [4] B. Rost, Enzyme function less conserved than anticipated, *Journal of Molecular Biology*.
520 318 (2002) 595–608. doi:10.1016/S0022-2836(02)00016-5.
- 521 [5] D.G. Naumoff, Hierarchical classification of glycoside hydrolases., *Biochemistry*.
522 *Biokhimiia*. 76 (2011) 622–635. doi:10.1134/S0006297911060022.
- 523 [6] D.G. Naumoff, GH97 is a new family of glycoside hydrolases, which is related to the α -
524 galactosidase superfamily., *BMC Genomics*. 6 (2005) 112. doi:10.1186/1471-2164-6-
525 112.
- 526 [7] T.M. Gloster, J.P. Turkenburg, J.R. Potts, B. Henrissat, G.J. Davies, Divergence of
527 catalytic mechanism within a glycosidase family provides insight into evolution of
528 carbohydrate metabolism by human gut flora., *Chem Biol*. 15 (2008) 1058–1067.
529 doi:10.1016/j.chembiol.2008.09.005.
- 530 [8] M. Kitamura, M. Okuyama, F. Tanzawa, H. Mori, Y. Kitago, N. Watanabe, A. Kimura,
531 I. Tanaka, M. Yao, Structural and functional analysis of a glycoside hydrolase family 97
532 enzyme from *Bacteroides thetaiotaomicron*., *J Biol Chem*. 283 (2008) 36328–36337.
533 doi:10.1074/jbc.M806115200.
- 534 [9] M. Okuyama, M. Kitamura, H. Hondoh, M.-S. Kang, H. Mori, A. Kimura, I. Tanaka, M.
535 Yao, Catalytic mechanism of retaining α -galactosidase belonging to glycoside hydrolase
536 family 97, *J Mol Biol*. 392 (2009) 1232–1241. doi:10.1016/j.jmb.2009.07.068.
- 537 [10] M. Okuyama, T. Yoshida, H. Hondoh, H. Mori, M. Yao, A. Kimura, Catalytic role of the
538 calcium ion in GH97 inverting glycoside hydrolase, *FEBS Lett*. 588 (2014) 3213–3217.
539 doi:10.1016/j.febslet.2014.07.002.
- 540 [11] L.E. Comstock, M.J. Coyne, *Bacteroides thetaiotaomicron*: a dynamic, niche-adapted
541 human symbiont., *Bioessays*. 25 (2003) 926–929. doi:10.1002/bies.10350.
- 542 [12] J. Xu, M.K. Bjursell, J. Himrod, S. Deng, L.K. Carmichael, H.C. Chiang, L. V Hooper,
543 J.I. Gordon, A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis.,
544 *Science*. 299 (2003) 2074–2076. doi:10.1126/science.1080029.

- 545 [13] A.A. Salyers, J.R. Vercellotti, S.E. West, T.D. Wilkins, Fermentation of mucin and plant
546 polysaccharides by strains of *Bacteroides* from the human colon., *Appl Environ*
547 *Microbiol.* 33 (1977) 319–322.
- 548 [14] F. Cuskin, E.C. Lowe, M.J. Temple, Y. Zhu, E.A. Cameron, N.A. Pudlo, N.T. Porter, K.
549 Urs, A.J. Thompson, A. Cartmell, A. Rogowski, B.S. Hamilton, R. Chen, T.J. Tolbert,
550 K. Piens, D. Bracke, W. Vervecken, Z. Hakki, G. Speciale, J.L. Munõz-Munõz, A. Day,
551 M.J. Peña, R. McLean, M.D. Suits, A.B. Boraston, T. Atherly, C.J. Ziemer, S.J.
552 Williams, G.J. Davies, D.W. Abbott, E.C. Martens, H.J. Gilbert, Human gut
553 *Bacteroidetes* can utilize yeast mannan through a selfish mechanism, *Nature.* 517 (2015)
554 165–169. doi:10.1038/nature13995.
- 555 [15] J.N. D’Elia, A.A. Salyers, Contribution of a neopullulanase, a pullulanase, and an α -
556 glucosidase to growth of *Bacteroides thetaiotaomicron* on starch., *J Bacteriol.* 178 (1996)
557 7173–7179.
- 558 [16] M. Okuyama, K. Matsunaga, K. Watanabe, K. Yamashita, T. Tagami, A. Kikuchi, M.
559 Ma, P. Klahan, H. Mori, M. Yao, A. Kimura, Efficient synthesis of α -galactosyl
560 oligosaccharides using a mutant *Bacteroides thetaiotaomicron* retaining α -galactosidase
561 (BtGH97b), *FEBS J.* 284 (2017) 766–783. doi:10.1111/febs.14018.
- 562 [17] W. Li, H. Fan, C. He, X. Zhang, X. Wang, J. Yuan, Z. Fang, W. Fang, Y. Xiao,
563 PspAG97A: A halophilic α -glucoside hydrolase with wide substrate specificity from
564 glycoside hydrolase family 97, *J Microbiol Biotechn.* 26 (2016) 1933–1942.
565 doi:10.4014/jmb.1606.06047.
- 566 [18] C. He, J. Li, W. Li, Y. Xue, Z. Fang, W. Fang, X. Zhang, X. Wang, Y. Xiao, Structures
567 of PspAG97A α -glucoside hydrolase reveal a novel mechanism for chloride induced
568 activation, *J Struct Biol.* 196 (2016) 426–436. doi:10.1016/j.jsb.2016.09.009.
- 569 [19] S. Yu, K. Bojsen, B. Svensson, J. Marcussen, α -1,4-Glucan lyases producing 1,5-
570 anhydro-d-fructose from starch and glycogen have sequence similarity to α -glucosidases,

- 571 BBA-Protein Struct M. 1433 (1999) 1–15. doi:10.1016/S0167-4838(99)00152-1.
- 572 [20] M. Okuyama, W. Saburi, H. Mori, A. Kimura, α -Glucosidases and α -1,4-glucan lyases:
 573 structures, functions, and physiological actions, Cellular and Molecular Life Sciences. 73
 574 (2016) 2727–2751. doi:10.1007/s00018-016-2247-5.
- 575 [21] J. Larsbrink, A. Izumi, G.R. Hemsworth, G.J. Davies, H. Brumer, Structural enzymology
 576 of *Cellvibrio japonicus* Agd31B protein reveals α -transglucosylase activity in glycoside
 577 hydrolase family 31, J Biol Chem. 287 (2012) 43288–43299.
 578 doi:10.1074/jbc.M112.416511.
- 579 [22] T. Miyazaki, Y. Ishizaki, M. Ichikawa, A. Nishikawa, T. Tonozuka, Structural and
 580 biochemical characterization of novel bacterial α -galactosidases belonging to glycoside
 581 hydrolase family 31, Biochem J. 469 (2015) 145–158. doi:10.1042/bj20150261.
- 582 [23] G. Speciale, Y. Jin, G.J. Davies, S.J. Williams, E.D. Goddard-Borger, YihQ is a
 583 sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids., Nat Chem
 584 Biol. 12 (2016) 215–217. doi:10.1038/nchembio.2023.
- 585 [24] T. Tagami, E. Miyano, J. Sadahiro, M. Okuyama, T. Iwasaki, A. Kimura, Two novel
 586 glycoside hydrolases responsible for the catabolism of cyclobis-(1→6)- α -nigerosyl, J
 587 Biol Chem. 291 (2016) 16438–16447. doi:10.1074/jbc.M116.727305.
- 588 [25] S.H. Light, L.A. Cahoon, K. V. Mahasenan, M. Lee, B. Boggess, A.S. Halavaty, S.
 589 Mobashery, N.E. Freitag, W.F. Anderson, Transferase versus hydrolase: The role of
 590 conformational flexibility in reaction specificity, Structure. 25 (2017) 295–304.
 591 doi:10.1016/j.str.2016.12.007.
- 592 [26] Y. Gozu, Y. Ishizaki, Y. Hosoyama, T. Miyazaki, A. Nishikawa, T. Tonozuka, A
 593 glycoside hydrolase family 31 dextranase with high transglucosylation activity from
 594 *Flavobacterium johnsoniae*, Biosci Biotechnol Biochem. 80 (2016) 1562–1567.
 595 doi:10.1080/09168451.2016.1182852.
- 596 [27] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H.

- McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega., *Mol Syst Biol.* 7 (2011) 539. doi:10.1038/msb.2011.75.
- [28] V. Alva, S.-Z. Nam, J. Söding, A.N. Lupas, S. I., C. S., C. S., F. M., G. A., G. L.J., The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis, *Nucleic Acids Res.* 44 (2016) W410–W415. doi:10.1093/nar/gkw348.
- [29] V. Lefort, J.-E. Longueville, O. Gascuel, SMS: Smart Model Selection in PhyML, *Mol Biol Evol.* 6 (2017) 461–464. doi:10.1093/molbev/msx149.
- [30] T.J. Wheeler, J. Clements, R.D. Finn, Skyline: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models., *BMC Bioinformatics.* 15 (2014) 7. doi:10.1186/1471-2105-15-7.
- [31] L. Kall, A. Krogh, E.L.L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server, *Nucleic Acids Res.* 35 (2007) W429–W432. doi:10.1093/nar/gkm256.
- [32] T. Goldberg, M. Hecht, T. Hamp, T. Karl, G. Yachdav, N. Ahmed, U. Altmann, P. Angerer, S. Ansorge, K. Balasz, M. Bernhofer, A. Betz, L. Cizmadija, K.T. Do, J. Gerke, R. Greil, V. Joerdens, M. Hastreiter, K. Hembach, M. Herzog, M. Kalemans, M. Kluge, A. Meier, H. Nasir, U. Neumaier, V. Prade, J. Reeb, A. Sorokoumov, I. Troshani, S. Vorberg, S. Waldrapp, J. Zierer, H. Nielsen, B. Rost, LocTree3 prediction of localization, *Nucleic Acids Res.* 42 (2014) W350–W355. doi:10.1093/nar/gku396.
- [33] W. Kabsch, XDS, *Acta Crystallogr D Biol Crystallogr.* 66 (2010) 125–132. doi:10.1107/s0907444909047337.
- [34] B.W. Matthews, Solvent content of protein crystals., *J Mol Biol.* 33 (1968) 491–497.
- [35] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, *J Appl Crystallogr.* 40 (2007) 658–674. doi:10.1107/s0021889807021206.

623 [36] P.D. Adams, P. V Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd,
624 L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R.
625 Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart,
626 PHENIX: a comprehensive Python-based system for macromolecular structure solution,
627 *Acta Crystallogr D Biol Crystallogr.* 66 (2010) 213–221.
628 doi:10.1107/s0907444909052925.

629 [37] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, *Acta*
630 *Crystallogr Sect. D Biol. Crystallogr.* 60 (2004) 2126–2132.

631 [38] V.B. Chen, W.B. Arendall 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J.
632 Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure
633 validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr.* 66
634 (2010) 12–21. doi:10.1107/s0907444909042073.

635 [39] Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the
636 TM-score, *Nucleic Acids Res.* 33 (2005) 2302–2309. doi:10.1093/nar/gki524.

637 [40] N. Terrapon, V. Lombard, H.J. Gilbert, B. Henrissat, Automatic prediction of
638 polysaccharide utilization loci in *Bacteroidetes* species, *Bioinformatics.* 31 (2015) 647–
639 655. doi:10.1093/bioinformatics/btu716.

640 [41] D. Ndeh, A. Rogowski, A. Cartmell, A.S. Luis, A. Baslé, J. Gray, I. Venditto, J. Briggs,
641 X. Zhang, A. Labourel, N. Terrapon, F. Buffetto, S. Nepogodiev, Y. Xiao, R.A. Field, Y.
642 Zhu, M.A. O'Neill, B.R. Urbanowicz, W.S. York, G.J. Davies, D.W. Abbott, M.-C.
643 Ralet, E.C. Martens, B. Henrissat, H.J. Gilbert, Complex pectin metabolism by gut
644 bacteria reveals novel catalytic functions, *Nature.* 544 (2017) 65–70.
645 doi:10.1038/nature21725.

646 [42] E.C. Martens, E.C. Lowe, H. Chiang, N.A. Pudlo, M. Wu, N.P. McNulty, D.W. Abbott,
647 B. Henrissat, H.J. Gilbert, D.N. Bolam, J.I. Gordon, Recognition and degradation of
648 plant cell wall polysaccharides by two human gut symbionts, *PLoS Biology.* 9 (2011).

doi:10.1371/journal.pbio.1001221.

- [43] S. Vidal, T. Doco, P. Williams, P. Pellerin, W.S. York, M.A. O'Neill, J. Glushka, A.G. Darvill, P. Albersheim, Structural characterization of the pectic polysaccharide rhamnogalacturonan II: evidence for the backbone location of the aceric acid-containing oligoglycosyl side chain, *Carbohydr Res Res.* 326 (2000) 277–294. doi:10.1016/S0008-6215(00)00036-7.
- [44] J.N. Glushka, M. Terrell, W.S. York, M.A. O'Neill, A. Gucwa, A.G. Darvill, P. Albersheim, J.H. Prestegard, Primary structure of the 2-O-methyl- α -l-fucose-containing side chain of the pectic polysaccharide, rhamnogalacturonan II, *Carbohydr Res Res.* 338 (2003) 341–352. doi:10.1016/S0008-6215(02)00461-5.
- [45] H. Ichinose, Z. Fujimoto, M. Honda, K. Harazono, Y. Nishimoto, A. Uzura, S. Kaneko, A β -L-Arabinopyranosidase from *Streptomyces avermitilis* is a novel member of glycoside hydrolase family 27., *J Biol Chem.* 284 (2009) 25097–25106. doi:10.1074/jbc.M109.022723.
- [46] S. Lansky, R. Salama, H. V. Solomon, H. Feinberg, H. Belrhali, Y. Shoham, G. Shoham, Structure–specificity relationships in Abp, a GH27 β -L-arabinopyranosidase from *Geobacillus stearothermophilus* T6, *Acta Crystallogr D Biol Crystallogr.* 70 (2014) 2994–3012. doi:10.1107/S139900471401863X.
- [47] S.-P. Nie, C. Wang, S.W. Cui, Q. Wang, M.-Y. Xie, G.O. Phillips, The core carbohydrate structure of *Acacia seyal* var. *seyal* (Gum arabic), *Food Hydrocolloids.* 32 (2013) 221–227. doi:10.1016/j.foodhyd.2012.12.027.
- [48] S.-P. Nie, C. Wang, S.W. Cui, Q. Wang, M.-Y. Xie, G.O. Phillips, A further amendment to the classical core structure of gum arabic (*Acacia senegal*), *Food Hydrocolloids.* 31 (2013) 42–48. doi:10.1016/j.foodhyd.2012.09.014.
- [49] L. Lopez-Torrez, M. Nigen, P. Williams, T. Doco, C. Sanchez, *Acacia senegal* vs. *Acacia seyal* gums - Part 1: Composition and structure of hyperbranched plant exudates,

675 Food Hydrocolloids. 51 (2015) 41–53. doi:10.1016/j.foodhyd.2015.04.019.

676 [50] B.B. Buchanan, W. Gruissem, R.L. Jones, eds., Biochemistry & Molecular Biology of

677 Plants, Second Edi, Wiley Blackwell, 2015.

678 [51] G.O. Aspinall, M.C. Knebl, The location of α -d-galactopyranose residues in gum arabic,

679 Carbohydr Res. 157 (1986) 257–260. doi:10.1016/0008-6215(86)85074-1.

680

Figure captions

Fig. 1. TLC analysis of BT3661 catalyzed hydrolysis of polysaccharides. Gum arabic from acacia (left two lanes) or arabinogalactan from larch wood (right two lanes) was incubated with (+) or without (–) BT3661 for 1 h at 35 °C in 40 mM sodium acetate buffer, pH 5.5, containing 10 mM CaCl_2 and 0.002 $\text{mg}\cdot\text{mL}^{-1}$ BSA. The sugars on the TLC were developed twice using a solvent system of acetonitrile:water (4:1, v/v). Standard arabinose and galactose are in the leftmost lane.

Fig. 2. Overall structure of the BT3661 monomer. The N-terminal domain (residues 20–300) is blue, the catalytic domain (residues 301–554) is magenta and the C-terminal domain (residues 555–647) is green. The catalytic residues (Asp408 and Glu463) and the calcium ion are shown by stick and sphere representations, respectively.

Fig. 3. Comparison of the active site structure of BT3661 with that of BT1871 (PDB ID: 5E1Q) (A, B) and multiple sequence alignments of GH97 retaining glycosidases (C). (A) Stereo view of a superposition of the active site of BT3661 (cyan) and the catalytic nucleophile mutant (D415G) of BT1871 (orange). For residues involved in ligand (GalLac; β -lactosyl α -galactoside, yellow) binding, the conserved residues are represented by stick model. The residue numbers are those of BT3661. (B) Stereo view of a superposition of the active site of BT3661 and BT1871-D415G. β -L-Arap (1 $\rightarrow$ 3) L-Araf (turquoise) was superimposed to GalLac. The differences are shown by stick model. The color codes are the same as (A). (C) Multiple sequence alignments around residue differences between BT3661 and BT1871 along with a phylogenetic tree. The alignments and the tree were constructed using amino acid sequences of the catalytic domain of the retaining GH97 enzymes. Similarity colors are based on the BLOSUM62 scoring matrix. The residues relevant to the C6 hydroxy group of α -galactoside are boxed in red. Secondary structure elements and residues numbers of BT1871 and BT3661 are depicted on the top and bottom, respectively.

707

708 **Fig. 4.** Graphical representation of the sequence conservation of amino acid residues in the region
709 surrounding Asn338 (left) and Gln378 (right) of the UniProt Reference Cluster that includes
710 sequences with at least 50% identity to BT3661 (UniRef50_F9D4D2). The residue numbers are
711 those of BT3661. Sequence logos were generated using Skyline (<http://skylign.org/>).

712

713 **Fig. 5.** Substrate-binding pockets of BT3661 (A) and BT1871 (B), and alignments of the
714 sequences forming the +1 subsite (C). Secondary structure elements and residues numbers on the
715 top and bottom of the alignment were those of BT1871 and BT3661, respectively. Similarity
716 colors are based on the BLOSUM62 scoring matrix. The residues relevant to the aromatic clamp
717 are boxed in red.

Table 1.

Summary of crystallization conditions, data collection and refinement statistics.

BT3661	
PDB ID	5XFM
Data collection	
Beamline	SPring-8 BL41XU
Space group	<i>P</i> 1
Unit cell parameters <i>a</i> , <i>b</i> , <i>c</i> (Å) α , β , γ (°)	72.8, 90.5, 128.6, 105.5, 94.8, 96.2
Wavelength (Å)	1.0
Resolution range (Å)	50.0–2.3 (2.38–2.3) ^a
<i>R</i> _{meas} (%)	15.0 (47.1)
$\langle I/\sigma(I) \rangle$	6.6 (2.7)
Completeness (%)	90.0 (81.9)
Redundancy	3.2 (3.2)
Refinement	
No. reflection	125,374
<i>R</i> _{work} / <i>R</i> _{free} (%)	20.7/25.3
No. of atoms	
Macromolecules	20,162
Ligand/ion	4
Water	1,257
<i>B</i> -factors (Å ²)	
Macromolecules	29.7
Ligand/ion	23.8
Water	28.8
Estimated coordinate error	0.31
RMSD from ideal	
Bond lengths (Å)	0.005
Bond angles (°)	1.03
Ramachandran plot	
Residues in favored regions (%)	95.6
Residues in allowed regions (%)	4.2
Residues in outliers regions (%)	0.2

^a Values in parentheses are for the highest resolution shell.

* $R_{\text{meas}} = \sum_{hkl} \{N(hkl) / [N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\langle I(hkl) \rangle$ and $N(hkl)$ are the mean intensity of a set of equivalent reflections and the multiplicity, respectively.

** $R_{\text{work}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$, *R*_{free} was calculated for 5% randomly selected test sets that were not used in the refinement.

Table 2.

Hydrolytic rates on 2 mM PNP glycosides.

Substrates	BT3661	BT3664
	v ($\mu\text{mol}/\text{min}/\text{mg}$)	
PNP-Ara	0.558	6.45×10^{-2}
PNP-Gal	0.208	3.17
PNP-Glc	N.D. ^a	1.03×10^{-3}
PNP-Xyl	N.D.	2.29×10^{-4}

^a N.D.: not detected.

Table 3.

Kinetic parameters for hydrolysis of PNP-Ara and PNP-Gal by BT3661 and its variants.

	PNP-Ara			PNP-Gal		
	k_{cat}	K_{m}	$k_{\text{cat}}/K_{\text{m}}$	k_{cat}	K_{m}	$k_{\text{cat}}/K_{\text{m}}$
	(s ⁻¹)	(mM)	(s ⁻¹ mM ⁻¹)	(s ⁻¹)	(mM)	(s ⁻¹ mM ⁻¹)
Wild type	5.51 ± 0.23	14 ± 1	0.400 ± 0.003	3.55 ± 0.30	25 ± 4	0.14 ± 0.01
N338E			0.510 ± 0.006 ^a			0.0154 ± 0.0001 ^a
N338A			0.0145 ± 0.0001 ^a	11.5 ± 0.22	11 ± 0	1.2 ± 0.0
N338D			0.199 ± 0.006 ^a	5.01 ± 0.23	11 ± 2	0.44 ± 0.05
Q378A	4.53 ± 0.20	11 ± 1	0.427 ± 0.013	3.34 ± 0.05	5.5 ± 0.4	0.61 ± 0.03
N338E/Q378A			0.419 ± 0.004 ^a	0.406 ± 0.031	24 ± 4	0.017 ± 0.002
N338D/Q378A			0.254 ± 0.003 ^a	3.39 ± 0.13	8.3 ± 1.3	0.42 ± 0.05

^aThese values were obtained from hydrolytic rates at low substrate concentrations of 1–2.5 mM.

Fig. 1. Kikuchi et al

Fig. 2. Kikuchi et al

Fig. 3. Kikuchi et al

Fig. 4 Kikuchi et al

Fig. 5. Kikuchi et al

Biochimie

Supplementary materials for

A novel glycoside hydrolase family 97 enzyme: bifunctional β -L-arabinopyranosidase/ α -galactosidase from *Bacteroides thetaiotaomicron*.

Asako Kikuchi^{a,†}, Masayuki Okuyama^{a,*,†}, Koji Kato^b, Shohei Osaki^a, Min Ma^a, Yuya Kumagai^a, Kana Matsunaga^a, Patcharapa Klahan^a, Takayoshi Tagami^a, Min Yao^b, Atsuo Kimura^{a,*}

^a Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan.
Kita-9 Nishi-9, Kita-ku, Sapporo 060-8589, Japan.

^b Faculty of Advance Life Science, Hokkaido University, Sapporo, Japan.
Kita-10 Nishi-8, Kita-ku, Sapporo 060-0810, Japan.

*Corresponding authors: M Okuyama, okuyama@abs.agr.hokudai.ac.jp;
A Kimura, kimura@abs.agr.hokudai.ac.jp.

Supplementary Fig. 1. Phylogenetic tree of GH97 members. Amino acid sequences of the catalytic domain (PF10566) were aligned by Clustal omega. A maximum likelihood phylogeny was estimated from the alignment obtained using PhyML-SMS. Amino acid sequences having Glu at β-strand 5, Asp at β-strand 4, and Asp at β-strand 5 were colored by green, blue, and yellow, respectively.

PDBID_3a24A

PDBID_3a24A

1 10 20 30 40

PDBID_3a24A R H M E L S P D G N L K T T I . T I . . G D R . L T Y D I T C N G R Q I L T P S P I S M T
 tr_A6L9E7 . A . K Q N Y Q V S P N R D L K V E V T V . A D K . V T F A I V Q D H S V M . N S A V S M T
 tr_R6WPH8 . A . K Q S Y Q L Q S P D K K L Q A V V T V . G D D . I R F S F T H D G T E V L A A S P I S M T
 tr_D4IM97 . A . A K T Y L L K S P D G K I A V R V D . G A D A . I T Y S I S R D G K E L I A P S P L S L M T
 tr_R6KSN5 I A . Q K Q Y L K S P N G K L E L S I . E A . D D Q . I C S L R H N T A I I V S S P I S M S
 tr_D6CY08 . A . K Q V Y E L K D I S G K V Q V N V . I V . N D K N . V E Y S V L H D N D V M A P S P I F M K
 tr_A5FA11 Q Q . K K N F S V S P N G K I E V T V . A V . S D K . I S W T I W H Q K D L I L A P S E M S M T
 tr_E4TIA6 Q K . T K D F L K S Q D G N I T L H V . V A . G K K . L W S V Q L K G K Q I I A P S A M S L Q
 tr_R7EPT5 A Q . A Q Q Y Q L N S P D G K L I V N I . E A . G Q T . L N W A I A H D G T T V L L P S D I A M Q
 tr_A6KZF1 A K . E K K Y L S P D G T L K V E I S A . G N E . L A Q V M H G M D T I L S H S N I G L V
 tr_A4ASM2 N T . A Q D L N V K S P D G K I T L N I . G I . S E T . I N W S A S L N G N I I E K A I V S M N
 tr_B1ZN91 A V . P S L P F L Q S P D G R I A V Q I R A . G G Q . L A D V L V D A T V V V R D A T L A L Q
 tr_C7PTT2 H A . Q Q R Y T V V S P D K T V T L T I . Q V . G A D . I T Y S L S Q D N K E L I S P S V V S F K
 tr_A3QFB6 A M . G E T Y Q L S P D G S I L V D V . I A . E E Q . I S Y R V S V D G Q T I L A P S H I A M T
 tr_Q5L8Y9 S V Y A K V I D V M P N D G A I K V S V D I . K D R . I Y S V S Y D N D Q L L K D C Y L N L
 tr_Q8A2L1 Q A . Q K N V E L S P N G E I K V S V R L . T D K . I Y D V A C N D E V L L K N Y L Q L K
 tr_D5BGL4 N A . Q Q Q Y L D S P D K D L K I N I . E I . S D S . I Y S Y V A Q G E L L G D N T L S L Q
 tr_D2QSA9 P R . P F Q R T S P N G R V A L T V S A . T D S . V R F A V R F A D Q S L V S P S A I A M T
 tr_C6W680 N A . Q A T R E V S P N G A I R L V A . Q L . G D Q . L R Y S V Y F N E I V I S P S E I D M V
 tr_R5UY73 . V . A K I I E A G S P N D K K N V I I . E T . D N Q . V S Y S L S R N G T K I L D T C P L S L T
 tr_E8V2F0 L A . Q K G L N L S P N G K L V M S F . S L . D A S V A G A S D G S G R . Q L Y S M I R G K S L I D A S A L G L E
 OWNDD_5xfma . . . L Q D V V V K G P D E K L Q L A S V F Q N E T K . P C Y S V S Y N G K T M L E K S P L G M N
 tr_A6K233 A F . A Q Q A T V T G P D S F L K V N V . S V K . Q G I . P V Y S V T Y K D K T I L E D S P L G F I
 tr_R5VXS7 A F . A Q Q A A V T G P D S R L K L D F . Q L Q . D G K . P V Y S V T Y D G K T V L E N S P L G F V
 tr_R5VWB3 L Y . A S E K E V L S P D G R L K V V V . S D E . G G I . P S Y Q V F Y D G V T F I E K S P L G M K
 tr_D6DQJ3 A S . A E N Q S I T S P D G K L V V T V . S D M . D G R . P S Y S V S Y D N V L F L K P S P L G M I
 tr_A5FD29 S L . N A Q S V I K S P D G K L V S L F V S . N G R . P F Y S I S Y N E K T F L E K S P L G L K
 tr_R5LVV6 S A . L T A Q E V Q G P D G A L K V N V . E L K . D G S . P V Y S V H Y N K I F L E P S P L G L K
 tr_R6B488 A Q . A N N Y T V T S P N G K L V M N V . K C E . G G K . A S Y T V D L N G K Q M L A A S A L G L V
 tr_R6C1F0 A Y . A E K N T V S S P D G K L N V V V . E D R . D G K . L Y Y S I D Y A G R K M M E S Q L G L L
 tr_A6KWX6 T . . M Q N N L L Q S P N G K I A I E Y . H A E K G F S . V I Y H S R D K A V M M S L P E I G L K
 tr_R5V624 T V . Q A G N D L S P N K K I G V S W D H P A K G L S . V I Y K G G D A S V P V V Q L K N I G L Q
 tr_Q8A1J7 S . . Q K D T G V I S P N G K I C L S V . E I K E D T E K D G F G K V F F N . V E Y K D G K T S R R V L S D T R I G L E
 tr_R6KK91 T A . P N A W E L V S K D S N V R F A L E N K Q E N G S T M . L S Y S V F Y K D G V A M E K S L L G L

PDBID_3a24A ▶ TT

β5 β6 β7

50 60 70 80

PDBID_3a24A LDNGT.....VWGENAKISGT.SRKSVDDEMIPSPF.YFASELRNHYNGLTRFKK....
tr_A6L9E7 LQGGE.....VLGANPKVLKV.TKTSVDKEIPSPF.YKQDVVKDIYNEMQISFRG....
tr_R6WPH8 LQNGV.....VLGAGPKVSKV.LKAAVDKVIPSPF.YKQTEVQDIYNEMTLSFRG....
tr_D4IM97 LADGR.....VLGRKASVRKA.SAKSVDDEIEAPF.YKRSAVEDRYNSLTTFKG....
tr_R6KSN5 LSNKN.....ILGVKSKVKNL.RRRVIDENITTAI.YKRSEIRNRCNELSLTFRG....
tr_D6CY08 LTDGT.....AFGLNPKVKN.SRRSVNETIYPPI.YKKKSICKQFNELTIDFKG....
tr_A5FA11 LDENT.....VLGKNPVVLNT.KKESVNTFETPL.YKKKTQONKYNQQLVINFKN....
tr_E4T1A6 LGGND.....VLGDNAKVSSA.KTEKINTTFKALN.YKKTNVQDVYNELTLTKCN....
tr_R7EPT5 CTEAEGKPIAFTFGKNVKVTKA.TRTSVDTSFPTPF.YKKANVKDTYNQLTLKCRN....
tr_A6KZF1 LENGT.....IVGKTPRITGE.RRRKIKDNIESPF.YRFKEFVATGNELDLKLGK....
tr_A4ASM2 MGNGR.....VIGDSPQLRNK.RFKSVNETLLVQVPNKDAMVQNNYNELSLNFRG....
tr_B1ZN91 VEQT.....TLGRPAELRTV.TPROHDATVQPVVRQKAATLRDRYHELRLLEFAG....
tr_C7PTT2 TDADK.....GNAWKVSKA.KTGAHDGVLTPVIMQKTKTIADRYNELHLDFSN....
tr_A3QFB6 FADGT.....ALGKSPQVVEV.KRSAISEQLTPVIKVR TAKIDNQYNQLSLSKFD....
tr_Q5L8Y9 LQNE.....TLGTNPHLRST.KRGTIDESVKREIPFKNAIVRNHCNTLRMNFSG....
tr_R8A2L1 LKDK.....ILGEHPKLSGQ.KQNKIKETLTPVVPLKFSTINNEYNLRLNFKG....
tr_D5BGL4 LKNE.....VLGVKPKIKKV.KNTSGTEERSPIVPLKFSNICKDFKAIETEFKG....
tr_D2QSA9 LGSGQ.....TLGAGVPQRS..RLRTVRDSILPPVVTKRRVIRDHYTELVLEFRD....
tr_C6W680 LTDGR.....KLSDGLARPOF.TERNVSETIISVPVEKRIRIPDRYKELKITLRN....
tr_R5UY73 VGGND.....TWGTD.KCRKI.TRKSVSEKVEFIVPRKYKETVDNYNQVELIYK....
tr_R5V2F0 LDDKA.....PLGSRVITDS.TPNQGVDDYALLF.TKVSHVHDPYNSLTVKEASNS
OWNID_5xfmA TNIG.....DFTKNLKLITGH.SVDKIDTVYQQTR.IKVSNVHYRANELTCHLENEQQQ
tr_A6KZ33 ANVG.....DFSRLMTFTGQ.KENKIDKTYTQDR.IKQSQIHYYQANELTCTFTNKEKK
tr_R5VXS7 SNIG.....DFSRLQMSFVGQ.QADKVEKTYTQDR.IKCSNVNYSANKLTVTLENAEKK
tr_R5VWB3 TTIG.....DFTSSLSLGSEIKETSIRKNYRLPN.IKKSVEYAAESVVVPTKEGRT
tr_D6D0J3 ANIG.....DFSSGMSLEKNVSTNKIDETVELAS.IKKSVRVYANEAVFSFTQQGKT
tr_A5FD29 TNVE.....DFTTGLALKTNPVQNKIDENYQLPN.IKNSNVHYEANEAVFSFTKENKA
tr_R5LYV6 TSIG.....DFSARLKQLGH.QVRTIDETYTMPH.AKVSRVHYQANELVCNLYLNEKGD
tr_R6B488 ANYG.....DFSQGLAMGAL.KGAPKHLTYVMDH.TKKSHEKDVYEATIGFNLAKKD
tr_R6C1F0 ANVG.....DFSQGLTYQGK.NEIKVEKSYDLRT.AKVSHVDYHANLNVVTYINGKKQ
tr_A6KWX6 TS.....DTDDCFKLISSTPVTSVDDYEMLT.GKRRHCHNEANECTYRFENVKGL
tr_R5V6Z4 LQP.....TSGDEFIDVSSSAVTPIREDYTMIT.GKRRHCTNEGTTQIFHFKNQONV
tr_Q8A1J7 TQLM.....SFTDNIRLKSVSQTKLEDDYVMLT.GKRSHCQNRGNERIIFENEKAE
tr_R6KK91 MDNC.....EYKDAQEVSASAVKDISNTYQLKS.GKMNLTDECREQETETRTKENK

PDBID_3a24A

β8 β9 β10 β11 β12

90 100 110 120 130

PDBID_3a24ADWNVEFRAYNDGIAYRFVNQG....KK..PFR...VVTQVSDYCFPSDMTASVP
tr_A6L9E7DYG LFRVYNDGVAYRF TTKR....KE..PIV...IADBEAVYNFPSDYKTFTA
tr_R6WPH8NYG LFRMYNDGLAYRF TTKM....KN..DIV...VVDDEADYNTFSSDHMAFAP
tr_D4IM97DYG LFRAYDDGAAYRF TTKM....KD..EIT...VTDLVDIRFPEDFTTLVP
tr_R6KSN5GFG LFRAYDEGVAYRF VVTE....NK..PFN...VIAEATFNFDDDMVYIIP
tr_D6CY08GYS LFRAYEDGAAYRF VSEL....KK..PFM...VESQASFCFPNDPKVFFVA
tr_A5FA11DFS LFRVYDDGAAYRF ITKK....KK..NIT...VKWENVSLNFQDQDNTLMP
tr_E4T1A6DFGVKFRVYNDGVAYRF FTKK....KG..EIV...IKNEANFNFTQDYKAFIMP
tr_R7EPT5GYS IQFRAYDDGAAYRF VSEL....RK..PFC...VKDEIAEFNFORDYHAFVFP
tr_A6KZF1GFG IIFRAYNEGVAYRF YTTQ....SS..DII...IKEEQAEFNFKEDYTAFLP
tr_A4ASM2EYQLLFRAFDDGIAYRF IDKS....SR..TKK...VVNBQMSLTFPEGTSSFFP
tr_B1ZN91NYAVVFRAYDEGVAYRF ETQLA....QS..EVK...VFAEAAFNFADNQTVWYP
tr_C7PTT2GLALEWRAYDNGIAWHWLTTS....KK..PYK...VLDLQAGFSLDKEGKAWYP
tr_A3QFB6NYQVDLFRAFDEGVAYRF SAQQ....SG..KSA...LTSQAEFNFARGAFAFYFP
tr_Q5L8Y9NYAVEFRVFDNGIAAYRF VTDK....KG..DNI...VMGEDFAINFPTNYKAHLS
tr_Q8A2L1NYSVEFRAFNDGIAYRF ITRQ....KG..EIE...VLGEDLSLQFPDYLHLHQ
tr_D5BGL4NYS IEFKAYNEGIAYRF KLGK....KG..SLE...VIOENFDIVFPAEYTLHLQ
tr_D2QSA9QHALLFRVYDDGVAYRF RTAF....TD..SIT...VRTETARFAFADNPTVYYP
tr_C6W680AFTVIIFRVYDDGVAYRIATSL....PG..EIT...VRSETAFFRFPANHPVYFP
tr_R5UY73DYKIEFRIYNDGVAYRF VSTA....NN..KQP...VKKESVSVFRFGQDHTSYTL
tr_E8V2F0 GGRNGRSMVVEARAYDTGIAFRYVLPS....QA..SITQLHLRGQTEFRFQSDDDTAWVL
OWNID_5xfmAKLGVIIFRVSDNDVAFRY TLPHQG...GK.ASVT...VKEEQTGFRFPEQTTTFLC
tr_A6KZ33NINIIIFRVSNNDIAFRY EMPKYG...DT.GSIV...IEKBTGFDFFPSFTTTFLC
tr_R5VXS7KLDIVFQLSNNDIAFRY EMPQYG...ET.ACMV...IEKEATGFDFFPSFTTTFLS
tr_R5VWB3AFDVEFRVSNNDVAFRY TVYPKK...DI.LCCV...IQOETTGFFVMPEGTTTFLC
tr_D6D0J3IYDVIIFRISNNDVAFKY KMYPQG...ET.LSCV...VKQEVTFGFAFPDGTTFFLC
tr_A5FD29AIDIIIFRVSNNAAPFKYKVYPQK...ES.RSCV...VEGEASGFLLPBGTTTFLC
tr_R5LYV6TLQVIIFRVSNNDIAQAY QISSPK...HT..HCI...IEKETTGFDFFPSYTTTFIT
tr_R6B488SMTLHLHVSNDVAYKY EMSRPKKDNP.KSVI...IYKEVSGFNFPEKTTTFLC
tr_R6C1F0PMTVTFNVSNNNDVAFRY TFDQL...KA.QHLI...VNEKTAFNLPKVTTTTYLC
tr_A6KWX6RVDLIFRVYNDGIAYRF HLPK....TK.EEIV...VLDLQYTAFAFTPGIKRWT
tr_R5V6Z4SLDLLELRVYDDGIAYRF VFPQ....SA.KEIC...VTDLQYTAFAFVEGQITRWT
tr_Q8A1J7TLDIVFRAYDDGIVFRYSLPS....VQ.DKDS...LTSBHTAYYIPEGTRKRWI
tr_R6KK91QFNLIIFRVYNDGVAFRYELPG....EDGQMHT...LQABHTEFAVPVNGKAWI

$\alpha 1$ $\beta 13$ $\eta 1$ $\beta 14$
 PDBID_3a24A 140 150 160 170 180
 PDBID_3a24A YVKS GK.DGDYNSQFFNS FENT YTT.DK TSKL.....NK..QRLMFL PLVV DADGDGVK V
 tr_A6L9E7 YVNS TK..PTFEEQFFNS FEQPY VN.ET ISSL.....ND..KRLKIL PLL VDLGDGRK V
 tr_R6WPH8 YVNS KK..ATFEEQFMNS FEQPY VH.EP ITKL.....NS..KRLMIL PLL VELDGGKK L
 tr_D4IM97 YARDGKKESGLARYYFNS FESTY ST.QP VSKV.....AD..NRLAFL PIL ADAG.TAK I
 tr_R6KSN5 YVRGGK.NKKVEDQFFNS FENIY TH.VNL SEM.....GT..NQLAFT PVV VELKDGGK L
 tr_D6CY08 SPKGRM.NEGKKDPFYSS FQNTY LE.TAL SAW.....DK..EQIAFL PVL VEGKNGKK I
 tr_A5FA11 YVRDLR...NPKDQFISS FESHY EN.KK ISGF.....SK..DTLAFS PFL IDYKNHKK A
 tr_E4T1A6 YMWDIR...DGKIFNSS FEALY KE.TK ISKF.....AT..DSLAFPL PLM VDEGDNKK V
 tr_R7EPT5 YINDNR...AGERYCYS FESY DE.AP LSRM.....YK..DSLAIT PLM VELENGKK A
 tr_A6KZ33 YTTNDK.....KPMAMA YQNVY DI.TP LSK.....AQ..PKLAFL PVT VDCG.SVK L
 tr_A4ASM2 Q.....EESMYSHNERD YLK.KD ISEM.....KN..GDFCSL PVM FNTK.NAK V
 tr_B1ZN91 E.....EESFFSHNERI FLR.RT LGDL.....AA..KNLASL PAV VEAG.QTR I
 tr_C7PTT2 Q.....EDAMFSHNERKY IQ.YK LDSI.....GE..KQLASL PAL FEVKG.GTK L
 tr_A3QFB6 F.....EKKFASATQPRF TP.IA AKGI.....DK..DELGSL PAL FVVN.GVN V
 tr_Q5L8Y9 Q.....PDGFKTS YECPY TH.VD TEKYA.....AT..DRMSYL PVL IETDKAYK I
 tr_Q8A2L1 Q.....PGGFKTA YEEPY TH.IS SHSWK.....PS..DKMSVL PVL IDTRKSYK I
 tr_D5BGL4 Q.....TPSFKTS YEPY TH.LK SSDFT.....AE..ESMSVL PVL IETTGNNHK I
 tr_D2QSA9 AVQPRP...EVDRFHTS FEEPY QI.RP LNQM.....PD..SVLCFS PVL LAPANGPR V
 tr_C6W680 EVVPPT...NMDIYHTG FEEPY TI.RP LENI.....SS..KNLFFS PVL VAPKEAPK I
 tr_R5UY73 L.....TDQLQNW FEQDY TV.KP INAL.....PK..DSFSVA PVM VEVD.KYK V
 tr_R5V2F0 A.....LPNRYSS YESEY LR.YN LSALS NQGGVAS..KFLIGF PTL LHQPGGAW M
 OWNID_5xfmA PQSDAM...IGWKRTKPS YEEY KADAP MS.D.....RSQYGHGYTF PCL FRIGNDGW V
 tr_A6KZ33 PQSDAM...IGWKRTKPS YEEY KADAP MN.V.....RSQYGHGYTF PCL FHVGENGW A
 tr_R5VXS7 PQSDPM...IGWMRTKPS YEEY VPDPE VA.T.....PSKYGEGYTF PAL FHVGDN.WA
 tr_R5VWB3 PQSGPM...GGFARTSPS YETPY TVDDT MG.....KNGWGEYTF PCL FKNSSDKGW I
 tr_D6D0J3 PQSKPM...GGFARTSPS YETSY TTDDV AG.....KNGWGEYTF PCL FRNGDNGW V
 tr_A5FD29 PQSKPM...TGFARTMPS YETSY TLDEP MG.....KNGIGEYTF PCL FKVNSNGW V
 tr_R5LYV6 PQAPAG...DGWMGTKPS YEEY TLDEA TG.T.....PSRYKLYTF PAL FHVGDNGW V
 tr_R6B488 PQITPM...AGWERTKPS YEEY TPDAQ MN.V.....KSQFGVGYTF PCL FKVGEDGW V
 tr_R6C1F0 PQSDPL...IGFARTKPS YEDY KVDPQ PLT.A.....KSGYGHGYTF PCL FKVGGDGW V
 tr_A6KWX6QKFTVG YEGFY WPN TDGV.....ADNEGREWSF PTL FQPSDSAF V
 tr_R5V6Z4QKLNQA YEDFY LKN KGRV.....QGYTAAQWGF PAL FEVADSVF T
 tr_Q8A1J7QNYSAQ YEGFY PLN.KTRA.....ERKDNMNGY PLL LEPADSLF V
 tr_R6KK91 PY.....DWNDRHKPS YEQYC RSEID LRS.....ECGHGRGWAF PML FNT.NGVW M

$\beta 15$ $\beta 16$ $\beta 17$ $\beta 18$ $\eta 2$ $\beta 19$
 PDBID_3a24A 190 200 210 220 230
 PDBID_3a24A CITE SDLE.NY PGLY LSAS.EGANRLSSMH APY PKRT..VQGGHNQLQMLVKEHED...Y
 tr_A6L9E7 CITE ADLE.DFP GMFL LNNE.TGKPS LKAVH APY PKVK..EQGGHNQLQMLVKEREN...Y
 tr_R6WPH8 CITE ADLE.DYP GMFL LNNS.TDKPVLKPI F ASY PKVK..KQGGHNQLQMLVEERED...Y
 tr_D4IM97 VITE ADLE.DYP GMY LYNP.TGGKALRGNF APY PAKE..VQGGHNELQLLVTERED...F
 tr_R6KSN5 CIAE SDVE.SYP GMFVRNA.NQSNS LKGVF APY PKET..VQGGHNKLQKLVTARE...Y
 tr_D6CY08 CITE ADLM.NYP GMY VKHG.EHGYSLDGI F AAYPKTI..VDEVRG.LKGVVKSREP...Y
 tr_A5FA11 VFLE ADLE.DYP GLFVTNN.KNKTGFESR SKYPTQE..TNGGFNYLNKILTERAD...Y
 tr_E4T1A6 VILE ADLE.DYP GMY LNLN.DTHKGFMGVY APY PLET..KMGGYNNMNVIPTRKAE...Y
 tr_R7EPT5 VIME AGLS.NYP GMFL TANPQMPHG VKAEF APY PQET..VIGGYDRNLNLPVTRKAD...V
 tr_A6KZ33 TLL ESDLE.AYP GMFVQSQ.QGKYGLKGVF APY PAKT..DFY PWRK.QEYVETITD...F
 tr_A4ASM2 LFT ESS LH.NYP GLFLERG..SNSTMNAV P KYVLNTVPNEASSPDRNQIITEEAD...Y
 tr_B1ZN91 AILE SDVE.DYP GLWL RGT..SGSGLTATF P PYPLKE..ELTGDR..DVRVTAAD...Y
 tr_C7PTT2 LLTE SGLF.NYAGMWL RGO..GAGKLKAVF P YYPKEK..KVTS DR..DEQVTAREN...Y
 tr_A3QFB6 LLTE TD LQ.SYP GLWL RLGK..GDGNVYGVH P FELDK.....DEQYTHDLG...I
 tr_Q5L8Y9 LI SEADLS.DYPCMF LKST..GKNGMQSIF P KAPLAF..GEDGDR..SLKITEEAD...Y
 tr_Q8A2L1 LI SEADLS.DYPCMF LKGN..GTNGISSIF P KVPLAF..GEDGDR..SLKIEKEAE...Y
 tr_D5BGL4 LI SEADLL.NY PAMFV KGT...GNGLKSVF P KNPIDF..EDEGDR..SMKILKRAD...Y
 tr_D2QSA9 VLTE SDLL.DYPCMF LTRN...GNGLTGIF APY PLTE..KLVA AEYSQYIVSKRAD...F
 tr_C6W680 GITE SDLE.DYPCMF L TGN..AEKALFGNF APY PAEE..TETKELYSTRYVSKRAN...F
 tr_R5UY73 LLA EANLY.NYP GMY LQPA...LESFHGIF ANYPDKE..VPYEGDN.KIYASTRKD...Y
 tr_E8V2F0 SLME ADLE.GNSSAYVTNP.....TGS WAGHML.....SVKLS PRWDDPAY
 OWNID_5xfmA LVSE TGVDSRY CGSRLSDV.SEGNLYTVAF PMAEE.....NNGNGT...V
 tr_A6KZ33 LI SE TGVDSKY CGSHLSDA.TADGLYTLAF PMP EE.....NNGNGT...A
 tr_R5VXS7 LVSE TGVRSLY CASHLSDA.TKDGLYSIAF PNPGE.....MNGLGS...S
 tr_R5VWB3 LI SE TGVNSY YCGSRLM GH..EDGLYTI GF P OEGE.....FNGNGT...I
 tr_D6D0J3 LVSE TGVNGGY CASRL LGH..KEGVYTI GF P OEGE.....ANGNGT...V
 tr_A5FD29 LI SE TGVDSGY CASRL IGH..EKGLYTI GF P MAGE.....NNGNGT...T
 tr_R5LYV6 LVSE TGVSSH YAGTKL SEG.TQDGLYTI TF P EKAE.....NNGVGD...A
 tr_R6B488 LVSE TGVSSAY P GSRLSDY.EPGKGYTVAF P QKGE.....NNGVGS...E
 tr_R6C1F0 LVSE TGT HGNYP GCHISDY.DAAKGYQIAF P MEKE.....ADTIGS...T
 tr_A6KWX6 LLTE ANI LRDN T GAYL TN.ADSSIYKIKL S DERLI.....
 tr_R5V6Z4 LI SE ANV AKGNC GSWL NNT.SSASSYKVEMA EPTN.....
 tr_Q8A1J7 LI TE ANI LRGH S GSRLLYNG.NDMNQYQVRM TDDKLA.....
 tr_R6KK91 MI TE AHL DGS YPA TH I DNA.GTDKAYKIR P EADE.....PIVPDA...V

N-terminal domain

catalytic domain

PDBID_3a24A	240	250	260	270	280	290
	β20	β21	β22	α2	α3	TT
PDBID_3a24A	I A K V D K P R N F	P W R I A V T T T D K D	L A A T N S Y L L G A P S . R M S	D L S W I K P G K		
tr_A6L9E7	I A K T S G T R S F	P W R A F I I S R N D K E	L A D C D M V Y R L A . S P S R V N	D I S W I K P G K		
tr_R6WPH8	I A K T S G T R A F	P W R V F I V S E N D K Q	L A D C D M V Y R L A . S P C R L Q	D V S W I K P G K		
tr_D4LM97	I A R T S G T R T F	P W R V F A I A G R D I Q	L A D N D M V Y R L A . A P S R V E	D V S W I K P G K		
tr_R6KSN5	I A K C S G R R S F	P W R I A I V S S D D K E	L L D N D M V Y K L A . A P S R L E	D T S W I K P G K		
tr_D6CY08	I A R V E G N T A F	P W R V M V I A K D D A E	L L C N D M V Y K L A . T P A Q F T	D F S W I K P G K		
tr_A5FA11	L V K T K G T R T F	P W R A I V I S E N D T A	L A N N D M V Q K L A . E P S K I K	D I S W I K P G K		
tr_E4T1A6	I A K T N G T R S F	P W R V V V I S Q S D K E	L L N N D M V Q K L A . S P S R I T	D A S W V K P G Q		
tr_R7EPT5	I S K P N N A P N G R P T G K Q T Y	P W R A V L V V T H D T Q	L A N N D M M Q R L A . P E C R I A	D T S W I K P G Q		
tr_A6KZF1	I S R S R G S R S Y	P W R V L A I T E K D T D	M P V N N L V Y A L A . S P N R I G	D T S W I K T G K		
tr_A4ASM2	I A K I S G K R S Y	P W R V F I I S D D D R T F	V E S N L V A Q L A . K P S K I P	N A S W I K P G Q		
tr_B1ZN91	L A V T R G S R T Y	P W R I L A V A H R D G E	L L T N P L V Y L L A . A P S R V A	D T S W I K P G K		
tr_C7PTT2	I A N I Q G P Q S F	P W R L V M I A R Q D G D	L L S N Q L V Y Q L A . P P A . T G	D Y S W V K P G K		
tr_A3QFB6	V S K S R S Y	P W R I L A I A R S D G E	L L D N Q M S Y T L A . E P S R I E	D T S W I K P G K		
tr_Q5L8Y9	I A K T D G K R S F	P W R M V V I S K E D K E	L I E N E M V Y N L S . A P C V L E	D Y S W I K P G Q		
tr_Q8A2L1	I A K T S G K R N F	P W R Y F V I T K E D K Q	L V E N T M T Y K L A . D K N Q L E	D V S W I K P G Q		
tr_D5BGL4	V A K I D G N R T L	P W R Y F V I S D K D T E	L I E N T M S A N L A . T P N V L K	D V E W I K P G Q		
tr_D2QSA9	I A R T T G K R T F	P W R V L V I G . T D R D	L P A S D M V Y R L A . T P S R I G	D A S W V K P G K		
tr_C6W680	L A K T K G T R A F	P W R V L I I A A Q D K D	L P A N D L V Y R L A . S P S R V A	D V S W V K P G K		
tr_R5UY73	L I P T C G K R V F	P W R V T G I F D N A S S	I L Q S E L I Y L L S E E K E	Q A A D Y T W V K P G K		
tr_E8V2F0	A V T G T L P H H S	A W R V L G V A D E P G K	L I E N N L Q T D L S . P E S R V K	D T S W I H P G K		
OWNID_5xfmA	A P A F A L P G A T	P W R T I T V G D H L K P	I V E T T V P W D V V . S P L Y E T	K H D . Y R F G R		
tr_R5VWB3	S P G L A L P G S T	P W R T I T V G E N L K P	I V E T T I P W D V V . E P L Y P T	E H T . Y Q M G R		
tr_R5VXS7	T P G L A L P G V T	P W R T I T V G S G L K P	I V E T T V P F D V V . E P L Y E P S	Q E . Y K F G R		
tr_R5VWB3	S P G I A L P G K T	P W R T I T L G S T L A P	I V E T T V P F D V V . E P L Y E P S	K D . Y Y T A E Y G C		
tr_D6D0J3	S P G I A L P G E T	P W R T I T V G K T L A P	I V E T T V P F D V V . K P L Y P A K	G E . Y T Y G R		
tr_A5FD29	A P G I P L I G E T	P W R T I T V G E T L N P	I V E T T I P F D L V . K P K Y E A S	K E . Y K Y T K		
tr_R5LYV6	T I A A S I P S L T S	W K T I T V G E T L K P	I V E T T S A Y D N V . R P L Y E P T	Q V . Y Q P G R		
tr_R6B488	F A G I P L P G E T	P W R T I T V G A S L A P	I V E T T I P Y D V V . E P L Y E A S	Q N . Y K P S R		
tr_R6C1F0	S A A F I L P G S T	P W R T I T V G S D L K P	L V E T T I P Y D V V . E P R F K A S	Q K . Y G T G K		
tr_A6KWX6	 C P S G W Y S	P W R V A I I G . T L A D	I V E S T I L V T D V S . E P C K I Q	D T Q W I K P G L		
tr_R5V6Z4	 V S G R W C S	P W R V A I I G . S L A D	V V E S T I L V T D V S . E P C R L S	D V S W I K P G V		
tr_Q8A1J7	 L G D S F T S	P W R V L M I G . S L S D	I V E S T I L V T D V S . E P N K I S	D S W I M P G S		
tr_R6KK91	E P V S E L P W F T	P W R A I I I G D D L N T	I F Q T Q M I S H L N . P A S V V N	D E S W I K A G R		

PDBID_3a24A	290	300	310	320	330	340
			α4	β24	TT	TT
PDBID_3a24A	V A W D W W N D W . N L D G . .	V D F V T G V N N P T Y K A Y I D F A S A N G I E Y V I I D E G W A V N L Q A D L M Q V				
tr_A6L9E7	V A W D W W N D W . N L Y G . .	V D F R A G I N N P T Y K Y Y I D F A A E H G I E Y V I I D E G W A V N L Q A D M M Q V				
tr_R6WPH8	V A W D W W N D W . N I Y D . .	V D F R A G I N N T E T Y K Y Y I D F A A E H G I E Y V I I D E G W S V N M A A D L M Q V				
tr_D4LM97	V A W D W W N T W . N L Y G . .	V D F K S G I N N E T Y K Y Y I D F A A D H G I E Y V I I D E G W A V N K K A D L F Q V				
tr_R6KSN5	V A W E W W N S C . G L S G . .	V D F K A G V N N V T Y K A Y I D F A S K H G I E Y V I I D E G W A V N L K A D L F Q V				
tr_D6CY08	V A W D W W N D W . N L Y N . .	V D F R A G I N N E T Y K Y Y I D F A S K F G I E Y V I I D E G W A V P G K A D L F E V				
tr_A5FA11	V A W D W W N D W . N I Y N . .	V D F K A G I N T Q T Y K Y Y I D F A S K N K V E Y V V I D E G W S V . . E T D I M K H				
tr_E4T1A6	A A W D W W N N W . N I T H . .	V D F K A G I N N T P T Y K Y Y I D F A S A N K I S Y I I I D E G W S V . . E G D L S K V				
tr_R7EPT5	V A W D W W N T C . N L T G . .	V D F K A G M N T P T Y K K Y I D F A A E N Q L E Y I I I D E G W S G K . E S L M E D I				
tr_A6KZF1	V A W D W W N D W . N L K G . .	V P F K A G I N M D T Y K Y Y I D F A S R N G L E F I V I D E G W Y D P K S G D M L T V				
tr_A4ASM2	V A W D W Y N A N . N I Y G . .	V D F E A G V N D E T Y K Y Y I D F A S N N G I E Y V I I D E G W T K S . T T E I L E D				
tr_B1ZN91	V A W D W W N A N . N V F G . .	V D F K S G V N T A T Y K Y Y I D F A A A N G L D Y V V I D E G W Y K . . L G N V L D V				
tr_C7PTT2	V Q W D W W H Y N . N I Y D . .	V D F R A G I N N E T Y K Y Y I D F A A K Y G I E Y V L I D E G W C D . . T R D L M K L				
tr_A3QFB6	I A W D W Y N E N . N L K G . .	V D F K A G I N N T E T Y Q Y A D F A A D F G I E N I L I D D G W S S . . Q E D V L N V				
tr_Q5L8Y9	V S W E W W H D A . R L Y G . .	V D F R S G F N M D S Y K Y Y I D F A S K F G I P Y I I M D E G W A K N . T R D P F T P				
tr_Q8A2L1	V S W E W W N G A . T P Y G Q D V T F K A G C	N L E T Y K Y Y I D F A S K F G I P Y I I M D E G W A K S . T R D P Y T P				
tr_D5BGL4	V S W E W W N G A . S P Y N . .	V D F V A G F N E E T Y K Y Y I D F A S E F G I P Y I I M D E G W A A S . T T D P Y T P				
tr_D2QSA9	C T D E W I T N I . N L Y N . .	V P F R A G V N T E T Y K Y Y V D F A K R F G F D R I L I D A G W S D . . N N D L F K I				
tr_C6W680	A T D E W I I T S . N I F N . .	V P F K S G I N T A T Y K Y Y I D F A K R F G F E R I M M D A G W S D . . N N D L F K I				
tr_R5UY73	V L W D W W N D R . N I Y H . .	I N F K S G I N T D T Y L Y L V D Y A A K H H I E Y V L I D E G W S A . . R N D L L T L				
tr_E8V2F0	A S W N W W V D N V N G K G . .	E S G N K E F T T E T M K Y Y V D F A A K S G F P Y F F L D A G W A . . S Q D I T K T				
OWNID_5xfmA	G T W S W I L W Q	D G S I N Y D D Q V R Y I D F A S A M G Y E Y A L I D N W D T R				
tr_A6KZ33	G T W S W I L W Q	D G S I N F D D Q K K Y V D L A A A M G Y E Y V L I D N W D T N				
tr_R5VXS7	G T W S W I M W Q	D P S I N Y D D Q I R F I D L A S E M G Y E Y T L I D G G W E K N				
tr_R5VWB3	S T W S W I I A G	D A S M N M D D Q K R Y V D F C A A M G Y R T V L V D A L W D T Q				
tr_D6D0J3	G S W S W I I G M	D G S T N Y K Q L R Y I D F S A A M G Y Q S V L V D A L W D K Q				
tr_A5FD29	G S W S W I I K M	D N N T T F P V Q K O Y I D F S A R M G Y E T I L V D A L W D T Q				
tr_R5LYV6	S T W S W I L W Q	D A S C N Y K D Q Q T F I D L A A A M G Y E Y I L I D A L W D K Q				
tr_R6B488	Y T W S W L I W Q	D G S I N Y D D Q V K M I D V A A A Q G Y E A I L V D A W D K Q				
tr_R6C1F0	Y A W S W L I W Q	D E S C N Y N D Q K L F I D L A A T M G Y E Y V L V D A L W D T Q				
tr_A6KWX6	V S W I Y W A Y N	H G S R D Y Q I V K K Y I D F A A D F D L P Y M L I D A E W D A M				
tr_R5V6Z4	V S W I Y W A Y N	H G S R D Y Q I V K K Y I D F A A D F H L P Y M L I D A E W D E M				
tr_Q8A1J7	A S W V Y W A H N	H G S K N F K I V K D Y I D L A V D M K W P Y S L I D A E W N E M				
tr_R6KK91	A S W S W W S N G	G T P R D Y K A Q L K Y V D L S A E M G W E Y M L I D A G W Q N M				

PDBID_3a24A TT α5 β25 α6 α7
 350 360 370 380

PDBID_3a24A	V	K	E	I	D	K	E	L	V	D	Y	A	A	S	K	N	V	G	T	I	L	W	A	G	Y	H	A	...	F	E	R	...	D	M	N	V	C	R	H	Y	A	E	M	G	V	K														
tr_A6L9E7	I	P	E	I	D	L	Q	E	L	V	D	Y	G	K	S	K	N	V	G	I	L	W	A	G	Y	W	A	...	F	A	R	...	D	M	N	V	V	K	H	Y	S	D	M	G	V	K														
tr_R6WPH8	I	P	E	I	D	I	P	E	L	V	N	Y	G	K	S	K	N	V	G	I	L	W	A	G	Y	H	A	...	F	D	R	...	D	M	E	K	V	V	K	H	Y	S	N	M	G	V	K													
tr_D4IM97	V	P	E	I	D	L	P	E	L	V	A	Y	G	K	E	R	G	V	G	I	V	L	W	A	G	F	Y	A	...	T	E	R	...	D	L	E	R	V	C	K	H	Y	S	E	M	G	I	K												
tr_R6KSN5	V	P	E	I	D	L	K	E	L	V	A	Y	A	D	S	K	H	V	G	L	I	L	W	A	G	Y	Y	A	...	F	E	R	...	D	L	E	R	V	C	K	H	Y	S	E	L	G	I	K												
tr_D6CY08	I	P	E	I	D	L	K	E	L	I	S	Y	A	K	S	K	N	V	D	L	I	L	W	A	G	Y	R	A	...	F	E	K	...	D	M	D	R	V	C	K	H	Y	A	A	M	G	I	K												
tr_A5FA11	N	P	N	V	D	L	E	A	L	I	A	Y	A	K	E	R	N	V	G	I	L	W	A	S	W	M	A	...	L	H	E	...	N	I	D	G	V	F	D	N	Y	A	K	L	G	V	K													
tr_E4T1A6	S	P	A	I	N	L	K	E	I	V	D	Y	G	K	Q	K	H	V	D	V	I	L	W	A	T	W	S	N	...	I	A	R	...	Q	M	D	E	I	F	P	O	Y	S	K	M	G	V	K												
tr_R7EPT5	S	P	E	I	N	L	E	L	V	A	Y	G	N	Q	K	G	I	G	I	L	W	A	S	R	N	L	T	G	G	D	...	I	S	P	...	A	V	E	Q	V	I	S	H	Y	A	Q	M	G	I	K										
tr_A6KZF1	I	P	E	L	D	L	P	E	L	I	A	Y	G	K	S	K	G	V	E	I	V	L	W	T	V	F	N	V	...	L	D	S	...	Q	L	E	A	A	C	K	K	Y	A	D	M	G	I	K												
tr_A4ASM2	N	E	N	I	D	V	P	G	L	I	K	Y	A	K	S	K	N	V	E	I	L	W	V	L	W	K	P	...	L	Y	E	...	D	P	E	K	I	L	S	L	Y	A	S	W	G	A	V													
tr_B1ZN91	V	P	E	I	D	V	P	E	L	V	A	Y	G	R	Q	K	N	V	S	I	V	L	W	V	V	W	K	T	...	L	E	D	...	Q	L	Q	P	A	L	D	Q	F	A	Q	W	G	V	R												
tr_C7PTT2	T	P	G	M	N	I	P	E	L	V	A	Y	A	H	S	Q	G	I	K	L	S	M	T	L	A	A	T	...	L	D	K	...	Q	L	D	M	A	L	D	K	F	R	E	W	G	V	K													
tr_A3QFB6	L	P	G	I	D	M	Q	A	I	I	A	H	A	K	A	R	G	V	G	V	Q	I	L	W	P	Y	S	A	...	L	D	K	...	N	L	E	E	A	F	K	L	Y	A	K	W	G	I	N												
tr_Q5L8Y9	N	P	T	I	N	L	T	E	L	I	K	Y	G	K	D	R	N	V	K	I	V	L	W	L	P	W	L	T	...	V	E	N	...	H	F	D	.	L	F	K	T	F	A	D	W	G	I	A												
tr_Q8A2L1	N	P	D	V	L	H	E	L	I	R	Y	G	K	E	K	N	V	G	I	V	L	W	L	T	W	L	T	...	V	E	K	...	N	F	D	.	L	F	K	T	F	N	E	W	G	I	K													
tr_D5BGL4	N	P	D	I	D	L	F	E	L	I	E	Y	G	K	K	R	N	V	K	I	V	L	W	L	T	W	L	T	...	V	E	H	...	H	F	D	.	L	F	E	K	F	A	E	W	G	I	A												
tr_D2QSA9	T	P	G	M	N	I	P	E	L	V	A	Y	A	H	S	Q	G	I	K	L	S	M	T	L	A	A	T	...	L	D	R	...	Q	L	E	P	A	M	K	Q	F	Q	S	W	G	V	D													
tr_C6W680	N	P	D	I	D	M	D	E	I	V	A	Y	A	K	S	K	G	V	G	I	S	M	T	L	A	L	T	...	L	D	K	...	Q	L	E	P	A	L	K	Q	F	S	K	W	G	V	D													
tr_R5UY73	N	P	D	V	D	I	P	R	I	C	E	Y	A	T	E	K	G	V	G	I	Q	L	W	S	K	W	V	N	...	I	M	R	...	Q	M	D	E	A	F	A	Q	F	S	K	W	G	V	K												
tr_E8V2F0	N	G	R	V	D	V	P	E	L	V	R	Y	A	A	A	R	N	V	K	V	I	L	W	L	Y	S	E	A	...	V	M	K	...	Q	M	K	E	A	F	P	V	Y	E	Q	W	G	V	A												
OWNID_5xfmA	I	G	H	Q	R	M	K	S	L	V	E	Y	A	R	D	K	G	V	E	L	F	L	W	Y	S	S	S	G	Y	W	N	D	I	E	Q	G	P	V	N	R	M	D	N	A	I	R	K	R	E	M	K	W	L	Q	S	L	G	V	K	
tr_A6KZ33	I	G	R	E	R	M	K	D	F	I	D	Y	A	H	S	K	N	V	D	V	L	W	Y	S	S	S	G	Y	W	N	D	I	V	Q	G	P	T	N	Y	M	D	N	P	I	R	K	K	E	M	K	W	L	H	N	I	G	V	K		
tr_R5VXS7	I	G	R	D	R	M	E	L	F	K	Y	A	H	Q	K	N	V	D	V	F	W	Y	N	S	N	G	Y	W	N	D	A	P	Q	D	A	K	Q	C	M	S	N	S	V	A	R	K	K	E	M	K	W	L	Q	K	G	V	K			
tr_D60J3	V	G	I	D	K	I	E	L	A	A	Y	A	K	Q	K	N	V	D	V	L	W	Y	N	S	N	G	Y	W	N	D	A	P	Q	G	P	R	G	K	M	S	N	I	E	R	R	K	E	M	K	W	M	Q	R	I	G	I	K			
tr_A5FD29	I	G	R	D	K	I	E	L	A	K	Y	G	K	D	K	G	V	A	L	Y	L	W	Y	N	S	N	G	Y	W	N	D	A	P	Q	T	P	R	G	I	M	D	N	A	I	A	R	R	K	E	M	K	W	M	Q	S	I	G	I	R	
tr_R5LYV6	I	G	K	D	K	I	A	E	L	A	Q	Y	G	T	E	K	G	V	G	L	Y	L	W	Y	N	S	N	G	Y	W	N	D	A	P	Q	G	P	R	G	M	D	N	S	I	R	R	K	E	M	A	W	L	K	E	I	G	I	K		
tr_R6B488	I	G	E	N	M	P	S	L	I	A	Y	A	Q	S	K	G	V	D	V	L	W	Y	N	S	N	G	N	W	N	D	A	P	Q	G	P	K	H	R	L	D	S	A	P	A	R	Q	K	E	M	A	W	M	K	T	L	G	V	K		
tr_R6C1F0	I	G	R	K	R	I	E	L	S	K	Y	A	K	S	K	K	V	S	L	M	L	W	Y	N	S	N	G	F	E	N	D	A	P	Q	T	P	R	Q	I	M	N	N	S	I	A	R	K	K	E	M	A	W	M	K	K	I	G	V	V	
tr_A6KWX6	I	G	R	D	K	I	A	E	L	S	K	Y	A	Q	S	K	N	V	S	L	M	L	W	Y	N	S	N	G	V	A	N	D	A	P	Q	G	P	R	G	I	M	D	N	I	V	A	R	K	K	E	M	A	W	M	K	S	I	G	I	K
tr_R5V6Z4	G	N	G	G	N	L	N	D	A	M	N	Y	C	K	E	K	N	V	T	P	L	I	W	Y	N	S	S	T	A	W	...	C	D	P	T	P	L	Y	R	L	N	T	P	E	V	R	D	K	E	F	E	W	L	K	E	I	G	I	Q	
tr_Q8A1J7	G	N	G	G	N	V	E	D	A	L	A	Y	A	K	Q	K	N	I	A	P	L	I	W	Y	N	S	S	T	A	W	...	C	A	F	G	P	L	Y	R	L	N	K	P	E	M	R	Q	K	E	F	A	W	L	E	S	M	G	V	K	
tr_R6KK91	G	N	G	G	T	E	D	V	V	K	Y	A	Q	Q	K	G	V	G	V	L	W	Y	H	S	G	A	G	R	E	K	D	S	V	S	T	H	R	L	M	S	D	P	E	L	R	R	A	E	M	K	R	I	S	E	V	G	V	K		

PDBID_3a24A β26 α8 β27 α9 β28
 390 400 410 420 430 440

PDBID_3a24A	G	F	K	V	D	F	M	D	R	D	D	Q	E	M	T	A	F	N	Y	R	A	A	E	M	C	A	K	Y	K	L	I	L	D	L	H	G	T	H	K	P	A	G	L	N	R	T	Y	P	N	V	L	N	E	B	G	V	N	G	L	E
tr_A6L9E7	G	F	K	V	D	F	M	D	R	D	D	Q	E	I	V	D	F	L	Y	R	G	A	E	T	C	A	K	Y	K	M	M	V	D	Y	H	G	I	C	K	P	T	G	L	O	R	T	Y	P	N	V	L	N	E	B	G	V	N	G	L	E
tr_R6WPH8	G	F	K	V	D	F	M	D	R	D	D	Q	E	I	V	D	F	L	Y	R	G	A	E	T	C	A	K	Y	K	M	M	V	D	Y	H	G	I	C	K	P	T	G	L	O	R	T	Y	P	N	V	L	N	E	B	G	V	H	G	L	E
tr_D4IM97	G	F	K	V	D	F	M	D	R	D	D	Q	K	I	A	D	F	T	Y	R	G	A	E	T	A	A	K	Y	R	L	F	L	D	M	H	G	Y	H	K	P	A	G	I	O	R	T	Y	P	N	V	L	N	E	B	G	V	F	G	L	E
tr_R6KSN5	G	F	K	V	D	F	M	D	R	D	D	Q	A	M	V	D	F	H	Y	R	G	A	E	I	A	A	K	Y	H	L	M	L	D	Y	H	G	T	Y	K	P	T	G	M	N	R	T	Y	P	N	V	L	N	E	B	G	V	H	G	L	E
tr_D6CY08	G	F	K	I	D	F	M	D	R	D	D	Q	V	V	E	F	N	R	K	A	A	E	T	G	A	K	Y	K	L	L	I	D	L	H	G	T	F	K	P	T	G	L	O	R	T	Y	P	N	T	I	N	E	B	G	V	H	G	L	E	
tr_A5FA11	G	F	K	V	D	F	I	D	R	D	D	A	K	M	V	N	S	Y	E	I	A	Q	K	A	A	N	H	K	L	I	I	D	F	H	G	Y	M	K	P	T	G	I	O	R	T	Y	P	N	I	L	N	E	B	G	V	K	G	L	E	
tr_E4T1A6	G	F	K	I	D	F	D	R	D	D	Q	V	V	A	S	T	Y	A	I	A	K	K	A	E	Y	H	L	M	V	D	Y	H	G	I	Y	K	P	T	G	I	O	K	T	Y	P	N	V	I	G	E	B	G	V	K	G	M	E			
tr_R7EPT5	G	F	K	I	D	F	D	R	D	D	Q	L	A	I	S	S	T	E	Q	L	A	C	A	A	K	H	H	L	L	D	L	H	G	L	.	K	P	F	C	I	Q	R	A	Y	P	N	I	L	N	E	B	G	V	K	G	L	E			
tr_A6KZF1	G	F	K	V	D	F	L	D	R	D	D	Q																																																

η3 α10 η4 α11 β29
 PDBID_3a24A 450 460 470 480 490
 PDBID_3a24A QMKW...SS.PSVDQVKYDVMIPFIRQVSGPMDYTQCAMRNASK.....GNYY..PC
 tr_A6L9E7 QMKW...SSEKEFDMVTYDVTIPFIRMIAGPMDYTQCGAMRNAAR.....GNHR..SV
 tr_R6WPH8 NMKW...VA.STYDMPLYDVTIPFVRMVAGPMDYTQCGAMRNAIR.....KNYA..PI
 tr_D4IM97 QMKW...TS.EKTDMTYDVTIPFIRMIAGPMDYTQCGAMRNATR.....RNFR..AV
 tr_R6KSN5 QLKWF...SGSENVQVTVYDVTIPFIRMIAGPVDYTQCGAMKNGNK.....RNFR..AV
 tr_D6CY08 EMKW...AE.PGTDQVIYDVTIPFIRMIAGPLDYTQCGAMNNVIM.....KNFH..AV
 tr_A5FA11 NNKW...TPNDDVPLYDTTIPFIRMMAGPMDYTQCGAMRNATK.....SEFK..PS
 tr_E4T1A6 NYKW...ATEDQPRYAVSIPFIRMMAGPMDYTQCGAMRNAIK.....ANYH..PV
 tr_R7EPT5 NSKWEPVNGAPLHDFPRYDVTAPYLRLAGPMDYTQCGAMVNTTR.....SLFR..SV
 tr_A6KZF1 EVKW...TD.IKNNMPLYDVTIPYIRMMAGPVDYTQCGAMRNATK.....ADWR..AM
 tr_A4ASM2 NNKW...SSAITPEHNVTLPFVRMAAGPFDYTPGAMINMNKIDYETGGSNYA..PL
 tr_B1ZN91 WNKW...STHITPEHTATLPPFIRMMMGPMDFTPGATLNANR.....TGFA..AN
 tr_C7PTT2 ISKW...EDWITPSHTTTIPFIRMAAGPMDFTPGGMQNVQR.....DSWS..AI
 tr_A3QFB6 NNKW...SEVTSTHNLHLPPFIRMIAGPLDYTPGAMANAQQ.....KNFN..KV
 tr_Q5L8Y9 QGGN...CKPENSILYLPFMRNAVGPMDFTPGSMISAQP.....EDNR..ST
 tr_Q8A2L1 QMGG...CKPENSILYLPFMRNAVGPMDYTQCGAMISMQP.....NIYR..SE
 tr_D5BGL4 QMGR...ATPDNSLFLPFMRNAVGPMDYTQCGAMISMQP.....EFYF..SR
 tr_D2QSA9 YNIF...STKATPDHDLVLPFIRMVSGPMDYEPGLLSNATK.....EQFR..PI
 tr_C6W680 YNMW...SAKATPEHNVLLAFIRMLGGPLDYEPGLMNNATP.....KGFH..PM
 tr_R5UY73 YNKF...SKRATVTHDVIIIPYLRMVWGPMDYTQCGAMLNAHP.....ETFY..EN
 tr_E8V2F0 NNKV...GRDSPVDRAVPAWRLLVGPMDYTAGGFNDQTE.....DSFV..AR
 OWNID_5xfmA NMVF...NQHFCDDEAFNTCLHPFIRNTVSGMEFGGCLLNKRLN.....RNND..GG
 tr_A6KZ33 NLIF...QQHFCDEEAFNACLHPFIRNTIGCMFEGGTFLNKRLN.....RNND..GG
 tr_R5VXS7 NLIIF...NQHFDDMEAYNACLHPFIRNTIGCMFEGGTLLNKRHN.....RTND..GG
 tr_R5VWB3 NMNF...SQGSCDAEAFNACIHPFVRNTIGSMDFGGSTLNKYYN.....AKNEL..GG
 tr_D6D0J3 NLHF...SQGSCDNEAFNATLHPFIRNTVSGMDFGGGSLNKYYN.....ADNAP..RG
 tr_A5FD29 NLHF...GQASCDKEAMNASIHPFIRNTVSGMDFGGGSLNKYYN.....SDNTPNKG
 tr_R5LYV6 NLVF...GQYHADKEAQRSTLYPFIRNAVAMDFGGVFFNKHFS.....KDGV..SG
 tr_R6B488 NVVF...TDYHAKKEAFEMTMHPFIRNAVSGFDWGGVMMNKYFS.....KDNK..SR
 tr_R6C1F0 NVVF...SDYHAKQEAFAELTMHPFCRNNAVAMDWGGTITMKNYLS.....KDNK..SR
 tr_A6KWX6 WYNN...KPTLTDNAAWHNCTLPFTRNVIGPMDYTQCTFTNSQH.....
 tr_R5V6Z4 WYNN...KETLTDNAAWHNCTLPFTRNVVGPMDYTQCTFTDSQH.....
 tr_Q8A1J7 WYNN...KPLITDRAAEHNVTLPFTRNVIGPMDYTQCTFTDSQH.....
 tr_R6KK91 TLGR...QERCDQAAKHNAATVPFTRNVVSGMDYTQCTFSDKVR.....QGVF...

catalytic domain C-terminal domain
 β30 α12 β31 α13 α14 β32
 PDBID_3a24A 500 510 520 530 540 550
 PDBID_3a24A YSEPMISQGTTRCRQLALYVVFESPFNMICDTPSNYMRRE.PE.STAFIAEIPTVWDESIIVLD
 tr_A6L9E7 NSEPMISQGTTRCRQLALYVVFESPFNMICDTPSNYMRRE.PE.CTEFISNIPITVWDETVSLA
 tr_R6WPH8 YTEPMISQGTTRCHQLALYVVFESPLNMLCDNPSNYNRE.PE.CTEFIANIPITVWEKTVALLD
 tr_D4IM97 GSEPMISQGTTRCRQLAEYVVFESPFNMICDPSNYMRRE.PE.CTEFIASVPTVWDETVGLD
 tr_R6KSN5 NEEPMISMGTTTRCRQLAEYVVFESPLNMLCDSPVLYERE.SE.CTSYISDIPTVWDETRALN
 tr_D6CY08 YTEPMISQGTTRCRQLALYTIFDSPINMLCDAPTNYLKE.EE.CTKFIAAIPTVWNQTLPIIS
 tr_A5FA11 HSNPMISQGTTRCHQLALYTIFEAPLQMMADSPATAFMKE.QE.STDFIAKVPTTFDETAALD
 tr_E4T1A6 SSAPMSMGTRCNQLAMYVVFESPLQMLADNPITYMKE.QE.CTDFITGVPTTFDETVALLD
 tr_R7EPT5 NDHPMSQGTTRVHQMAMYTVEFAPLQMLADSPSKYMKKE.QE.CTDFIAKVPTTFDETVIALD
 tr_A6KZF1 YYTPASMGTRCHQLAAIYVHDSPFTMLCDAPTNYLNE.QE.CVDFIASLPVEVDSFTFIAS
 tr_A4ASM2 FTRPMAFGTTRAHQVAMYVVFESPLQMLCESPTIYYKE.QE.TVDFITQIPTVWDETVVLE
 tr_B1ZN91 FARPMISQGTTRCHQLALYVVFESPLQMLCDSPSHYLRE.PD.MMEFLRSVPTVWDETVRLD
 tr_C7PTT2 PAEPLTLGSTRSNQLAMYVVFESPLQMLCDLPTHYHHE.PV.AMDFLQTVPTVWQQTIPQLQ
 tr_A3QFB6 WSRPMSLTTRAQQLAMYVIYESPLQMLADTPTSYRAE.PE.IPRFIAAIPITVWDDFKVLLH
 tr_Q5L8Y9 RANAMGSGTRAFQMALFIIFESGLQMLADNPVYYRE.LP.CTEFITSVPTVWDETVKVLV
 tr_Q8A2L1 RPNSASIGTRAYQMALFVVFESGLQMLADNPITYYRN.ED.CTRFITQVPTVWDETVVLE
 tr_D5BGL4 RPNSAAIGTRAQQMAFFVVFESGLQMLADNPITYYRN.ED.CTRFITQVPTVWDETVVLE
 tr_D2QSA9 SGQVMSQGTTRCHQLAMFLVYDSPISIFSGNPSQALLE.PR.FMEFLGSVPSAFDETVKLD
 tr_C6W680 PENVMISLTTRTQQLAMFVVFESPIQFFAGNPSQGLKE.PG.FMELLGSIPTVWDDTTRILD
 tr_R5UY73 QHEPMISQGTTRSHQLAMYVVFESPLQMLISDSTKYDKN.LR.SFEFISKIPITVWDETVPLE
 tr_E8V2F0 NATPQVMGTRAQQLALYVVFENPIPMVSDSPQNYDGE.AAPAFQFLKDVPTVWDETVHIG
 OWNID_5xfmA ...TTRRTTDFVQLATTVLLQNPVQNFALAPNNLKDVPAV.CMDFMKRVPTVWDETVRFVD
 tr_A6KZ33 ...NIRKTTDVFQLATAVLFQNPVQNFALAPNNLTDAQV.CLDFMKQVPTVWDETVRFID
 tr_R5VXS7 ...TIRKTTDVFQLATAVLYQNPVQNFALAPNNLTDAQV.CLDFMKQVPTVWDETVRFID
 tr_R5VWB3 ...RMRVTSDFVQLATAVLFQSPVQHAFALAPNNLTDAQV.AVDFMKKVPTVWDETVKFD
 tr_D6D0J3 ...SRRVTSDFVQLATAVLFQSPVQHAFALAPNNLTDAQV.AVDFMKKVPTVWDETVKFD
 tr_A5FD29 ...SKRITSDVFQLATAVLFQSPVQHAFALAPNNLTDAQV.AVDFMKKVPTVWDETVKFD
 tr_R5LYV6 ...TLRKTTDAFQLATSVLVQSGIQHFGITPNNLTEQPEF.ILDPMKKVPTVWDETVRFID
 tr_R6B488 ...HQRNTSDIFEMATAVTNQSSVNCVCLYPNNLQDVPAV.ELDWLKAAPTAVDDVKFIA
 tr_R6C1F0 ...HRRYTTDVFEMASAIMNQAAIQCIATYPNNLSELPQH.EIDFLKSVPITVWDETVKFD
 tr_A6KWX6 ...PHITSDAHEALALLVIFESALQHILADRPGEYRQOPVE.VQNFIRYLPVAVWEDTTRLLS
 tr_R5V6Z4 ...PHITSDGHEALALLVIFESALQHILADRPGEYHQAQPAE.IRSFISLPTVWDDTTRLLS
 tr_Q8A1J7 ...PHITTCGHEALALLVIFESALQHILADRPGEYSSLPKRVREFLSGLPTAVDDTTRLLC
 tr_R6KK91 ...ALRKTTMGHQLALA VVFESGFQCFADKAE SYLSLPEQ.PKQFLKEVPAAWDESIVLLA

PDBID_3a24A	→	β33	TT	β34	β35	T.T
	560	570	580	590		
PDBID_3a24A	GKMGEYIVTARR	KGD.VVYVVGITDWS	ARDTEVD	CS.FLGDK	SYHA	
tr_A6L9E7	GKVSSEYVAIARR	HGN.DWYVIGALT	NWT.PREL	LDLS.FLGE	G.DYTL	
tr_R6WPH8	GKVGGEFVAIARL	HGN.DWYVIGALT	NWD.AREVEL	LDLS.FLGD	G.NYKL	
tr_D4IM97	GRVGEYVSIARR	KGD.TWYVVGMT	DWN.ARTLT	IDLGL.FLGE	G.AYTA	
tr_R6KSN5	GKIGEYISMAR	KGD.VWYVIGALT	NWK.KRTME	FDFS.FLGE	G.KWTI	
tr_D6CY08	GKVGGEHIVMARE	KDG.IWYVVGGL	TDWK.ERDEV	VDLS.FLGD	G.EFNA	
tr_A5FA11	GEVGKYVSIARR	KGN.TWYVIGALT	NWD.SRDIT	IDDFS.FLEK	G.KKFQA	
tr_E4T1A6	GEVGEYAAALARQ	KDG.IWYVIGALT	NWN.SRQL	TLIDFS.FLPE	G.NYKA	
tr_R7EPT5	GKIAEYITIAARR	KGN.TWYVIGALT	NWT.PRSCT	IDLGL.FLSE	G.NYEA	
tr_A6KZF1	GELGKYIVTVRK	KDV.NWYVIGMT	NWD.ERDV	QLDDFS.FLPE	G.MSYTA	
tr_A4ASM2	GKVSDEYIIVARR	KGN.NWYVIGALT	DWT.SREFE	VDLS.FLNG	G.EFDM	
tr_B1ZN91	ARISDYVVVARR	RGA.DWYVIGALT	DWN.TRELE	LDLS.FLSE	G.RFTL	
tr_C7PTT2	AKVGEYVAIARQ	APNGDWYVIGALT	NWT.PREL	TLIDLS.FLGE	G.SFKI	
tr_A3QFB6	ASIGEYFTIARR	AGD.NWYVIGALT	NTQ.ARSQ	QINLD.FLGD	G.EYEM	
tr_Q5L8Y9	AKVGEAVVVAKR	KGE.QWYVIGIT	GNQ.PQNI	EIDLGL.FIP	AGQSFT	L
tr_Q8A2L1	AKVGEYVIVAKR	KGE.KWYVIGMT	NDKENEREFE	INLD.FLKD	GRTYRV	
tr_D5BGL4	AEVGEYVAIVAKR	KGK.QWYVIG	IANSKEDE	RTFKIKLD.FLPE	GETLOL	
tr_D2QSA9	AQLGQYIVSARQ	KGN.DWYVIGL	GNWS.AHD	LTLSLD.FLGD	G.AYEA	
tr_C6W680	ARLGDYVVVTARK	KGS.DWYVIGALT	NDS.ARELS	IPLD.FLPA	G.QFNA	
tr_R5UY73	GEVGEYIALARR	HND.TWYVIGV	INGAT.PRH	EIDLGL.FIGN	G.PKKI	
tr_E8V2F0	GEVGEYSTIVRR	NGR.EWYVIGS	MTNWT.PRD	LSVLS.FLGS	G.RYVA	
OWNID_5xfmA	GYPGKYVVLARR	QGD.TWYVIAA	VNAGKE.PLK	LKLDLE.M.FAG	.KTV	
tr_A6KZ33	GYPGKYVVLARR	HAD.KWYVIA	INAQKE.PLK	LKLNLP.MCTK	G.DKV	
tr_R5VXS7	GYPGKYCVLARR	HGA.RWYVVG	VNAQKE.PLK	LSLNLPMWEK	G.ETV	
tr_R6VWB3	GYPGKYIVMARK	HAG.KWYVAG	INAQDE.TIK	LKIDVSTLFAA	G.EIV	
tr_D6D0J3	GYPGKYVILARR	HGD.KWYVIA	VNAQKE.TLK	LKVNLP.MFSN	G.EKV	
tr_A5FD29	GYPGKYVILARR	KGT.KWYVIA	VNAQKE.TLS	LKLNLP.MMVSD	.ASL	
tr_R5LYV6	GYPGKYCVVARR	YGN.QWYVIA	AATNATKQ.PMK	LNLSLP.WLTN	.QQV	
tr_R6B488	GYPPTQYAVVARK	ASAANTGTAAA	SAG.RWYVVG	LNATDK.PLAL	TLNLP.MFAG	.KTV
tr_R6C1F0	GYPGKYAVVARR	HGD.KWYVAG	LNGTDQ.VMK	LTNLP.MLAG	.KSV	
tr_A6KWX6	GYPAESVVIARK	SKD.SWYVIG	LNGTEN.TKS	FRVNLFIKDS	.IQLI	
tr_R5V6Z4	GYPSSYVVLARK	SGD.CWYVIA	LNGTNE.KKS	LAVDLS.FIQE	P.VKEI	
tr_Q8A1J7	GYPGADIVLARR	KGD.VWYVIG	LNGTNE.NTL	LSFSLV.PLQVE	G.KTM	
tr_R6KK91	GYPADYAVVARR	SGD.VWYVIG	ISGKEE.ERE	TEFTLP.AGCE	G.KSF	

C-terminal domain

PDBID_3a24A	β36	β37	TT	β38	TT	β39
	600	610	620	630	640	
PDBID_3a24A	TLFKDGVN	AHRRAGR	.DYKCESFP	IKKDGK	LKVHLAPG	GGFALKIK
tr_A6L9E7	ELFKDGIN	ADRAAR	.DYKKVEVIP	VPTDRK	LKHMAPG	GGFYAARIQ
tr_R6WPH8	ELFKDGIN	ADRAAC	.DYKKVEVIP	VPADRK	VKKMAPG	GGWAAKIY
tr_D4IM97	ELFRDGVN	ADKAAR	.DYKRESVE	IPADRK	LVKKMAPG	GGFALKVS
tr_R6KSN5	ELYKDGIN	ADKIAS	.DYRKSVIF	VQNRKL	TVNLAE	GGGCAMKIY
tr_D6CY08	EIFRDGIN	ADRVGK	.DYKREVIR	VSDNKK	LNLMAPG	GGFVVRIT
tr_A5FA11	EIFSDGLN	ADKAAAT	.DYKKEIVT	VDSTTK	LKYRLANG	GGGLAMIK
tr_E4T1A6	VVFSDGIN	ADRDAT	.DYKKETKS	IKSGDK	LKIQAPG	GGWAARIE
tr_R7EPT5	EIFADGIN	ADREAT	.DYKKEVRI	VTAKDK	LVQLAPG	GGWTARIT
tr_A6KZF1	VLFKDGVN	ANKQAE	.DYRKETIR	IDKDSRL	TLHLASG	GGFAMKLE
tr_A4ASM2	QSYKDGIN	ASRNAQ	.DYKVSSST	VNENTK	LKIDLS	SGGYSAIFT
tr_B1ZN91	DEYRDGVN	ADRFQA	.DYKRSQVP	VNRETK	LKLRLAE	GGGYAARIS
tr_C7PTT2	DSWKDGIN	ADRNAK	.DCKRESGT	TDSHSK	LTIKMTKG	GGYVARIR
tr_A3QFB6	QLLQDGIN	AGNFAE	.DYKLIKKT	VTKGDS	LDLNMAAN	GGYAIQLI
tr_Q5L8Y9	TSFEDGIN	ADRQAM	.DYKKKEST	VNNQTRM	TLKMVRN	GGWAGTIK
tr_Q8A2L1	TSFEDGIN	AGYQAM	.DYRMKSAA	MNKNQK	LSVKMARN	GGFSAVLE
tr_D5BGL4	TSFEDGVN	AKRQAM	.DYVKRTKT	VKKDDV	LEITIVKN	GGWAAKIE
tr_D2QSA9	VSCDGVN	ADRNPA	.DYQFSTKT	VTRTESL	LPVHLAPG	GGFLIRLR
tr_C6W680	TICEDGVN	ADRYPA	.DYRISEAV	LTSRDL	LTIRLAPG	GGYMVRLR
tr_R5UY73	KAHIDGVN	AGQQA	.DSQIIEQV	IKDNEK	LPIDMSRG	GGYTGIIH
tr_E8V2F0	EIYEDGAD	AATSPK	.HVVHIRKET	VTAKST	LKIKLAPG	GGVAIHFI
OWNID_5xfmA	ALYKDDKK	G.....	EPELTS	SLKVKENGK	VQLEIRPQ	GGILC IK
tr_A6KZ33	MLYQDDKK	L.....	NPHAEI	ITLKKNDE	VSIIMQAE	GGFLLVK
tr_R5VXS7	SYLYLNDKK	R.....	QPQLEEL	IKNPAE	VKVVIQPD	GGIILTK
tr_R5VWB3	NLYEDNDK	L.....	AGSVKEQV	INKKQT	LNLSIPKN	GGVITRK
tr_D6D0J3	RLYFSDDK	L.....	QGGVKQIE	IGKKQEL	LQLAIPCN	GGVLIITK
tr_A5FD29	TCYFDDEK	L.....	NGKVKT	VHLKKNKE	TEIKIPAN	GGILIVN
tr_R5LYV6	SIIHDNEAR		TAGLSQGA	VNKKGI	LTIEMQPM	GGAMIYTK
tr_R6B488	TYLTDQPK	KKKGEKFF	.TSVKKTLK	VGKDGK	AKVVIQPN	GGIIE
tr_R6C1F0	TYYHETAS	KDKKALW	PVSQMKNVK	VDKKGN	LKVEMQPK	GGLIVEK
tr_A6KWX6	QLYFSDT	TDGK.....	QIKIEN	.SAFDTKE	LNIECQPR	GGFVWVK
tr_R5V6Z4	TLYADGKERN		QFDIQMLS	GLKKSVE	VDCLPR	GGVVMKVQ
tr_Q8A1J7	DVFKDGVDDK		SFAIEDI	QLSNDKM	DMSCLP	GGFVAVIK
tr_R6KK91	TMILIDGK	KDSF.....	DYMPVENT	I...GT	VKKVLPN	GGFAGMIR

PDBID_3a24A

```
PDBID_3a24A .....
tr_A6L9E7 .....
tr_R6WPH8 .....
tr_D4IM97 .....
tr_R6KSN5 .....
tr_D6CY08 .....
tr_A5FA11 .....
tr_E4T1A6 .....
tr_R7EPT5 .....
tr_A6KZF1 KNIPSFYQKYIETEGLYVTSSGKVSDEALLKACDIIISLMLAKRPDVKAHMOVKKGCHVMI
tr_A4ASM2 .....
tr_B1ZN91 .....
tr_C7PTT2 .....
tr_A3QFB6 .....
tr_Q5L8Y9 .....
tr_Q8A2L1 .....
tr_D5BGL4 .....
tr_D2QSA9 .....
tr_C6W680 .....
tr_R5UY73 .....
tr_E8V2F0 .....
OWNID_5xfmA .....
tr_A6KZ33 .....
tr_R5VXS7 .....
tr_R5VWB3 .....
tr_D6D0J3 .....
tr_A5FD29 .....
tr_R5LYV6 .....
tr_R6B488 .....
tr_R6C1F0 .....
tr_A6KWX6 .....
tr_R5V6Z4 .....
tr_Q8A1J7 .....
tr_R6KK91 .....
```

PDBID_3a24A

```
PDBID_3a24A .....
tr_A6L9E7 .....
tr_R6WPH8 .....
tr_D4IM97 .....
tr_R6KSN5 .....
tr_D6CY08 .....
tr_A5FA11 .....
tr_E4T1A6 .....
tr_R7EPT5 .....
tr_A6KZF1 GKDEETCDLPEFAHICNCEDSIKYWNWRARGFGGAPEDFSSSCGEENLLALPQDKYVGE
tr_A4ASM2 .....
tr_B1ZN91 .....
tr_C7PTT2 .....
tr_A3QFB6 .....
tr_Q5L8Y9 .....
tr_Q8A2L1 .....
tr_D5BGL4 .....
tr_D2QSA9 .....
tr_C6W680 .....
tr_R5UY73 .....
tr_E8V2F0 .....
OWNID_5xfmA .....
tr_A6KZ33 .....
tr_R5VXS7 .....
tr_R5VWB3 .....
tr_D6D0J3 .....
tr_A5FD29 .....
tr_R5LYV6 .....
tr_R6B488 .....
tr_R6C1F0 .....
tr_A6KWX6 .....
tr_R5V6Z4 .....
tr_Q8A1J7 .....
tr_R6KK91 .....
```

PDBID_3a24A

```

PDBID_3a24A .....
tr_A6L9E7 .....
tr_R6WPH8 .....
tr_D4IM97 .....
tr_R6KSN5 .....
tr_D6CY08 .....
tr_A5FA11 .....
tr_E4T1A6 .....
tr_R7EPT5 .....
tr_A6KZF1 N I L I H E F A H L I H T V G I V G V E P D F N E R L E A L R Q N A I R K G L W E K T Y A V S N K E E Y F A E C V Q S F
tr_A4ASM2 .....
tr_B1ZN91 .....
tr_C7PTT2 .....
tr_A3QFB6 .....
tr_Q5L8Y9 .....
tr_Q8A2L1 .....
tr_D5BGL4 .....
tr_D2QSA9 .....
tr_C6W680 .....
tr_R5UY73 .....
tr_E8V2F0 .....
OWNID_5xfmA .....
tr_A6KZ33 .....
tr_R5VXS7 .....
tr_R5VWB3 .....
tr_D6D0J3 .....
tr_A5FD29 .....
tr_R5LYV6 .....
tr_R6B488 .....
tr_R6C1F0 .....
tr_A6KWX6 .....
tr_R5V6Z4 .....
tr_Q8A1J7 .....
tr_R6KK91 .....

```

PDBID_3a24A

```

PDBID_3a24A .....
tr_A6L9E7 ..... KR.
tr_R6WPH8 ..... K.
tr_D4IM97 ..... RK.
tr_R6KSN5 ..... KK.
tr_D6CY08 ..... K.
tr_A5FA11 .....
tr_E4T1A6 ..... K I H
tr_R7EPT5 ..... KK.
tr_A6KZF1 F N C N R Y A E P A N G V H N W V N R R T K L K T Y D P D M Y R L L Q E Y F Y E I E I P I H N V V H E
tr_A4ASM2 ..... KK.
tr_B1ZN91 ..... AE.
tr_C7PTT2 ..... KQ.
tr_A3QFB6 ..... KR.
tr_Q5L8Y9 ..... MK.
tr_Q8A2L1 .....
tr_D5BGL4 .....
tr_D2QSA9 ..... KK.
tr_C6W680 ..... R.
tr_R5UY73 ..... I E K
tr_E8V2F0 ..... P A K
OWNID_5xfmA .....
tr_A6KZ33 .....
tr_R5VXS7 .....
tr_R5VWB3 .....
tr_D6D0J3 .....
tr_A5FD29 .....
tr_R5LYV6 .....
tr_R6B488 .....
tr_R6C1F0 ..... L R K
tr_A6KWX6 ..... MKN
tr_R5V6Z4 .....
tr_Q8A1J7 .....
tr_R6KK91 .....

```

Supplementary Fig. 2. Multiple sequence alignment of GH97 retaining enzymes. An alignment of full-length sequences and three-dimensional structures were made using MAFFTash (<https://sysimm.ifrec.osaka-u.ac.jp/MAFFTash/>). The alignment was rendered by ESPript 3.0 (<http://esprict.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>).