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_{cat}/K_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_{cat}/K_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

Abbreviations

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

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

77 phylum. *Bacteroides thetaiotaomicron*, which is known to be involved in the breakdown of plant
78 polysaccharides and host-derived glycans in the colon [11–13], possesses 10 GH97 proteins.
79 Among them, one inverting α-glucoside hydrolase [7,8] and a small number of retaining α-
80 galactosidases [9,14] have been characterized. The inverting enzyme, categorized into subfamily
81 97a, is a member of the starch utilization system (*sus*) and catalyzes the hydrolysis of α-glucosidic
82 linkages of maltooligosaccharides, which are produced by the action of SusA and SusG [8,15].
83 This enzyme has several appellations, including 97A1_BACTH, SusB, *Bt*GH97a, and BT3703,
84 which can cause confusion. We use the most commonly used locus-tag name in this paper. The
85 two retaining α-galactosidases, BT1871 and BT2620, which are members of subfamily 97b, are
86 believed to be involved in the degradation of the α-galactosidic linkage in galactomannan.
87 Substrate specificity of BT1871 has been examined for its transglycosylation properties [16].
88 BT1871 uses *p*-nitrophenyl α- and β-mannosides as accepter molecules and has a subsite structure
89 to accommodate α-galactosyl and α/β-mannosyl moieties. BT2620 forms part of a polysaccharide
90 utilization locus (PUL; MAN-PUL1) and has been proposed to degrade a terminal α-galactosidic
91 linkage in yeast α-mannan [14]. Apart from Bacteroidetes, a halophilic GH97 inverting α-
92 glucoside hydrolase from the Arctic marine bacterium (*Pseudoalteromonas* sp. K8) has been
93 found to prefer the α-1,6-glucosidic linkage to the α-1,4-glucosidic linkage, which is distinct from
94 BT3703 [17]. The tolerance to high concentrations of salts is due to a chloride-ion binding site
95 where a unique Arg residue plays a central role [18]. Only these enzymes have well characterized
96 structures and functions, with most GH97 proteins uncharacterized.
97 Novel specificities are frequently discovered within existing GHs. Among the evolutionarily
98 related GHs to GH97, namely GH27, 31 and 36 [6–8,16], GH31 is such a family that consists of
99 enzymes with distinct functions. The family comprises not only hydrolases and transglycosylases
100 but also an α-1,4-glucan lyase [19,20]. In addition, several enzymes with unique specificity,
101 including 1,4-α-glucan 4-α-glucosyltransferase [21], α-galactosidase [22], sulfoquinovosidase
102 [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′-ATGAAGAAACTTGTAAGCACT-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; *Sac*I-site underlined) and 5′-GCCCGGCTCGAGCTACTTTATACAAAGAATACC-3′ (antisense; *Xho*I-site underlined). The PCR product was restricted with *Sac*I and *Xho*I, 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; *Sac*I-site underlined) and 5′-GCGAAGCTTTTATTACTTTATTACTGCCAC-3′ (antisense; *Hin*dIII-site underlined). The resultant PCR product was restricted with *Sac*I and *Hin*dIII, 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

155 medium (10 g·L^{-1} Bacto tryptone, 5 g·L^{-1} Bacto yeast extract and 5 g·L^{-1} NaCl) supplemented
156 with 0.05 mg·mL^{-1} ampicillin at 37 °C to mid-exponential phase (OD_{600} = ~0.7). The culture was
157 cooled on ice for 30 min, at which time isopropyl β-D-thiogalactopyranoside was added to a final
158 concentration of 0.2 mM and cells were incubated at 12 °C for induction of BT3661 or at 15 °C
159 for BT3664. The induction temperature was optimal for each recombinant protein production.
160 After 24 h with vigorous shaking, the cells were collected by centrifugation (5,000 × *g*, 5 min,
161 4°C) and resuspended in 20 mM Hepes-NaOH, 300 mM NaOH, pH 7.5 (buffer-A), followed by
162 disruption through sonication. The supernatant was obtained by centrifugation (20,000 × *g*, 20
163 min, 4°C).
164 Recombinant proteins were purified from the above supernatant, loaded onto a Ni^{2+}-chelating
165 Sepharose Fast Flow column (GE healthcare, Chicago, Il, USA), which was prepared according
166 to the supplier's protocol and pre-equilibrated with buffer-A. After washing the column with
167 buffer-A containing 50 mM imidazole, the absorbed proteins were eluted with a linear gradient of
168 50 to 300 mM imidazole in buffer-A. Active fractions were passed through a Sephadex G-25 (GE
169 Healthcare) column pre-equilibrated with 50 mM sodium acetate, 100 mM NaCl and 25 mM
170 $CaCl_2$, pH 5.5 (buffer-B). The obtained fractions were pooled and concentrated using a 50 kDa
171 MWCO Vivaspin 20 (Sartorius, Göttingen, Germany). For preparation of protein crystallization,
172 the protein was further subjected to a Sephacryl S-200 (GE Healthcare) gel-filtration column pre-
173 equilibrated with buffer-B. The active fractions were pooled, dialyzed against 20 mM sodium
174 acetate buffer containing 10 mM $CaCl_2$, pH 5.5, and concentrated using a 50 kDa MWCO
175 Vivaspin 20. Fractions containing BT3661 were determined by specific activity and SDS-PAGE.
176 All purification steps were performed at 4 °C. The concentration of the purified protein was
177 estimated by amino acid analysis of a protein hydrolysate (6 M HCl, 110 °C, 24 h) using a JLC-
178 500/V (JOEL, Tokyo, Japan) equipped with a ninhydrin-detection system. The purification
179 procedures for BT1871 were described in our previous report [9].
180

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 $CaCl_2$ 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 Na_2CO_3/50 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 $CaCl_2$ 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 $CaCl_2$ (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 $mg \cdot mL^{-1}$) 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 $CaCl_2$ and 0.002 $mg \cdot mL^{-1}$ 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 F_{254} (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 $mL \cdot min^{-1}$. 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.

233

234 *2.6. Site-directed mutagenesis*

235 Point mutations to BT3661 were introduced by site-directed mutagenesis using the methods
236 for the PrimeSTAR mutagenesis basal kit (Takara Bio). The oligonucleotides used for each
237 mutation were as follows: for N338E, 5′-ATCGAT**GAA**TGGTGGGATACTAGAATC-3′ (sense;
238 substituted codon in boldface) and 5′-CCACCA**TTC**ATCGATCAACGCATACTC-3′ (antisense);
239 for N338A, 5′-ATCGAT**GCC**TGGTGGGATACTAGAATC-3′ (sense) and 5′-
240 CCACCA**GGC**ATCGATCAACGCATACTC-3′ (antisense); for N338D,
241 ATCGAT**GAC**TGGTGGGATACTAGAATC-3′ (sense) and 5′-
242 CCACCA**GTC**ATCGATCAACGCATACTC-3′ (antisense); for Q378A, 5′-
243 ATTGAA**GCG**GGACCGGTGAACCGAATG-3′ (sense) and 5′-
244 CGGTCC**CGC**TTCAATGTCATTCCAATA-3′ (antisense). The mutant enzymes were prepared
245 from *E. coli* transformants harboring each expression plasmid, followed by the aforementioned
246 purification procedure.

247

248 *2.7. Crystallization and data collection*

249 Crystallization screening was performed using a series of crystallization kits from Qiagen
250 (Hilden, Germany) by the sitting-drop vapor-diffusion method, in which 0.75 μL of 10 $mg \cdot mL^{-1}$
251 protein solution (20 mM sodium acetate containing 10 mM $CaCl_2$, pH 5.5) was mixed with an
252 equal volume of reservoir solution. The appropriate crystals of BT3661 for structural analysis
253 were obtained within 2 weeks at 20 °C under a reservoir solution containing 0.2 M lithium sulfate,
254 0.1 M Mes-NaOH buffer (pH 6.0) and 45% (v/v) 2-methyl-2,4-pentanediol.

255 Because the crystallization conditions of BT3661 contained 45% (v/v) 2-methyl-2,4-
256 pentanediol, which can be used as a cryoprotectant, the crystals were directly selected from
257 crystallization drops and flash cooled for X-ray diffraction experiments. Diffraction data were
258 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 $Å^3$ 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

285 enzymes were predicted by Phobious. BT3661 and BT3664 were predicted to have a secretion
286 signal sequence composed of 19 and 27 residues at the N-terminus.
287
288 *3.2. Production of BT3664*
289 Recombinant BT3664 was successfully produced in *E. coli* transformed with pCold-*bt3664*.
290 Substrate specificity of the recombinant enzyme purified by Ni^{2+}-affinity chromatography was
291 examined with 2 mM PNP-Gal, PNP-Ara, PNP-Glc and PNP-Xyl (Table 2). The enzyme
292 exhibited substantial activity toward PNP-Gal, which is ~50 times larger than that toward PNP-
293 Ara, and minimal activity towards PNP-Glc and PNP-Xyl. No product was detected on the TLC
294 plate from the reaction mixtures containing BT3664 and each polysaccharide (polygalacturonic
295 acid or pectin). These results indicate that BT3664 is an α-galactosidase.
296
297 *3.3. Production and characterization of BT3661*
298 Recombinant BT3661 was produced using the expression plasmid pCold-*bt3661*, which
299 encodes mature BT3661 with an N-terminal $(His)_6$-tag. The hydrolytic activities toward 2 mM
300 PNP glycosides were determined (Table 2). The enzyme hydrolyzed PNP-Ara and PNP-Gal with
301 velocities of 0.558 and 0.208 μmol/min/mg, respectively, but no activity was observed towards
302 PNP-Glc or PNP-Xyl. The optimum pH for the hydrolysis reaction was 5.5. After treatment at
303 various pH values for 24 h at 4 °C, BT3661 maintained more than 90% of its hydrolytic activity
304 when compared with that of the original enzyme over the pH range of 4–12. In the presence of 1
305 mM EDTA·2Na, the enzyme lost catalytic activity and this was restored by the addition of 5 mM
306 $CaCl_2$. Kinetic parameters of BT3661 for hydrolysis of PNP-Ara and PNP-Gal were determined
307 (Table 3). The k_{cat}/K_m value of BT3661 for PNP-Ara was 2.8 times greater than that for PNP-Gal.
308 The kinetic parameters of BT1871 toward PNP-Ara, which exhibited a k_{cat}/K_m value of 778 s^{-1}
309 mM^{-1} for PNP-Gal [9], were determined to know the hydrolytic activity of an unequivocal α-
310 galactosidase toward β-L-arabinopyranoside. The constants were 15.9 ± 1.8 mM and 68.9 ± 5.1

311 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
312 that for PNP-Gal. BT3661 exhibited the distinct substrate specificity of BT1871 and we thus
313 regarded BT3661 as a bifunctional β-L-arabinopyranosidase/α-D-galactopyranosidase.
314 The hydrolytic reaction of BT3661 on polysaccharide substrates, gum arabic from the acacia
315 tree and arabinogalactan from larch wood, was investigated. The release of arabinose and
316 galactose from gum arabic and that of arabinose from arabinogalactan were observed by TLC
317 (Fig. 1). Release rates of these monosaccharides were calculated through HPAEC-PAD analysis.
318 The rates of arabinose and galactose release from gum arabic were 2.02 and 1.43 μmol/min/mg
319 ($v/[E]_0$ = 2.47 and 1.75 s^{-1}, respectively). The release rate of arabinose from arabinogalactan was
320 40.9 μmol/min/mg ($v/[E]_0$ = 50.0 s^{-1}), but that of galactose was below the detection limit.
321
322 *3.4. Crystal structure of BT3661*
323 The crystal structure of BT3661 was determined at 2.3 Å resolution by the molecular
324 replacement method using the structure of BT1871 (PDB ID: 3A24 [3]) as the search model. The
325 structure was refined to crystallographic R_{work} and R_{free} values of 20.7% and 25.3%, respectively.
326 The model showed well-defined geometric parameters based on the Ramachandran plot (favored
327 region: 95.6%, allowed region: 4.2%, outliers region: 0.2%). The crystal contained four BT3661
328 molecules in the asymmetric unit (residues 16–647 for chain A, residues 20–647 for chain B,
329 residues 16–647 for chain C and residues 19–647 for chain D were built, respectively). The
330 structures of the four molecules were very similar with RMSDs (root mean square deviations) of
331 0.13–0.21 Å among 628–632 Cα atoms of respective chains. As the electron density for chain A
332 was more clearly visible than that for the other molecules, we have described the structure based
333 on chain A. The BT3661 monomer structure contains three domains (Fig. 2), which is similar to
334 other GH97 members: the $(\beta/\alpha)_8$ barrel-fold catalytic domain (domain A, residues 301–554) is
335 located between two β-sandwich folded domains, the N-terminal domain (domain N, residues 20–
336 300) 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 β→α 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-

363 hairpin of BT1871 and the β→α loop 3 of BT3661, is extremely low, and the pairs, Glu351–
364 His245 (BT1871) and Asn338–Gln378 (BT3661), are not common in GH97 retaining
365 glycosidases (Fig. 3C). All sequences that have Asn equivalent to Asn338 in UniRef50_F9D4D2
366 have Gln corresponding to Gln378 (Fig. 4).

367 Glu179 and Phe467 were placed near the +1 subsite, which is similar to BT1871 (Fig. 5A). In
368 addition, Trp171 is involved in the formation of the +1 subsite and forms a clamp-like subsite
369 structure together with Phe467. An aromatic residue equivalent to Trp171 was conserved in only
370 sequences that had a shorter evolutionary distance to BT3661 (Fig. 5C). BT1871 has no
371 corresponding tryptophan residue and no such aromatic clamp (Fig. 5B). In Fig. 3B and 5A, β-L-
372 Ara*p* (1→3) L-Ara*f*, which should be a part of a natural substrate of BT3661, was virtually placed
373 in the active-site pocket. The L-Ara*f* moiety appears to be stabilized by the aromatic clamp. While
374 the plus subsites of BT1871 are mainly composed of Trp474 of the β→α loop 6 of the catalytic
375 domain (Fig. 5B), those of BT3661 had a slit-like pocket that is narrow because part of a long
376 loop (Pro163–Glu180) extends from the domain N and the β→α loop 6 (Fig. 5A). A multiple
377 sequence alignment indicates a low conservation of the relevant region (Pro163–Glu180) (Fig.
378 5C), suggesting the divergence of the structure of the plus subsites in GH97 retaining enzymes.

379

380 *3.5. Substrate specificity of mutant BT3661s*

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

389 also prepared. The latter two double-mutant enzymes were prepared to ensure that the carboxy
390 group of Glu338 or Asp338 was in the same orientation as that of Glu351 of BT1871. The
391 optimum pH values of these mutant enzymes were 5.5–5.7, and almost identical to the wild-type
392 BT3661. Kinetic parameters of the mutant enzymes for hydrolysis of PNP-Ara and PNP-Gal were
393 determined (Table 3). The parameters for several mutant enzymes were not determined because
394 the K_m values were beyond the solubility limit of the substrates; however, the specific constants,
395 k_{cat}/K_m, were estimated at low substrate concentrations. Substitution of Asn338 with Glu affected
396 the hydrolytic activity toward PNP-Gal more significantly than to PNP-Ara. The K_m value of
397 N338E for PNP-Gal was beyond the measurement limit and its k_{cat}/K_m exhibited 10% that of the
398 wild-type enzyme. The K_m value for PNP-Ara was also greater than the measurable limit but its
399 k_{cat}/K_m was almost equivalent to that of the wild-type enzyme, indicating an increase in both k_{cat}
400 and K_m. The N338A mutation substantially reduced the hydrolytic activity toward β-L-
401 arabinopyranoside, but an increase toward α-galactoside was observed, resulting in a decrease in
402 the k_{cat}/K_m value by 3.6% and an increase by 860%, respectively. The mutation of Asn338 to Asp
403 gave a slightly negative effect on β-L-arabinopyranosidase activity, but exhibited a positive effect
404 on α-galactosidase activity. Replacement of Gln378 for Ala had negligible effect on the hydrolytic
405 activity towards PNP-Ara, but a 4-fold rise in the k_{cat}/K_m value was observed for PNP-Gal. The
406 double mutant enzymes, N338E/Q378A and N338D/Q378A, had very similar k_{cat}/K_m values
407 towards the substrates as the two corresponding single mutations, N338E and N338D.
408
409 **4. Discussion**
410 To find an enzyme with novel activity in GH97, we have focused on two GH97 proteins,
411 BT3661 and BT3664, which are categorized into a different subfamily, 97c, from function-known
412 sequences [6]. BT3664 was identified to be an α-galactosidase in this study. A gene encoding
413 BT3664 is located in RG-II PUL 3 (PUL64) together with genes encoding GH43_34 endo-1,4-β-
414 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→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

441 distinct molecular evolution from BT3664.

442 Since BT3661 exhibits the dual substrate specificity, which is found in GH97 for the first time,
443 and to understand its substrate recognition mechanism, its crystal structure was determined and
444 compared with the structure of BT1871, which strictly recognizes α-galactosidase. While both
445 enzymes have similar active-site pocket structures (Fig. 3A), different structural features were
446 found around a hydroxymethyl group of α-galactopyranoside (Fig. 3B). BT1871 stabilizes a C6
447 hydroxy group of the α-galactoside through a hydrogen bond with a Glu residue (Glu351) [16],
448 whereas BT3661 possesses Asn338 instead of this Glu residue. The mutational study suggests
449 that the bifunctional substrate specificity of BT3661 relies on the length of the side chain of
450 Asn338 and van der Waals interactions. Replacement of Asn338 by Glu, which was performed to
451 mimic BT1871, substantially decreases the specificity toward α-galactoside; although, the
452 negative charge at this location has a positive effect on its specificity, as shown by the N338D
453 mutation (discussed below). The Ala mutation increases specificity to α-galactoside. These results
454 indicate that a longer side chain at position Asn338 prevents binding of α-galactoside, whereas a
455 shorter side chain creates additional space within the substrate pocket to accommodate the 6-
456 CH_2OH of α-galactoside. Similarly, a substitution of Gln378 for Ala that increases the k_{cat}/K_m
457 value 4-fold seems to afford space for the hydroxymethyl group at C5. N338A displayed a
458 decrease in activity toward PNP-Ara, of which the k_{cat}/K_m value for hydrolysis was reduced to
459 3.6% that of the wild-type enzyme. A reduction in the size of the side chain may facilitate weaker
460 binding of β-L-arabinopyranoside because of the loss of a van der Waals contact. Substitution of
461 Asn338 for Asp showed 50% reduction in the k_{cat}/K_m value for hydrolysis of PNP-Ara. The size
462 of the side chain is near identical to that of Asn, but the negative charge of the carboxy group may
463 destabilize the interaction with L-arabinopyranoside. The substitution results yielded an increase
464 in the k_{cat}/K_m value for hydrolysis of PNP-Gal. The carboxy group of the Asp residue possibly
465 forms a hydrogen bond with the OH-6 and promotes charge separation ($O6^{\delta-}\cdots H6^{\delta+}$). Electron
466 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.

[13] A.A. Salyers, J.R. Vercellotti, S.E. West, T.D. Wilkins, Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon., Appl Environ Microbiol. 33 (1977) 319–322.

[14] F. Cuskin, E.C. Lowe, M.J. Temple, Y. Zhu, E.A. Cameron, N.A. Pudlo, N.T. Porter, K. Urs, A.J. Thompson, A. Cartmell, A. Rogowski, B.S. Hamilton, R. Chen, T.J. Tolbert, K. Piens, D. Bracke, W. Vervecken, Z. Hakki, G. Speciale, J.L. Munōz-Munōz, A. Day, M.J. Peña, R. McLean, M.D. Suits, A.B. Boraston, T. Atherly, C.J. Ziemer, S.J. Williams, G.J. Davies, D.W. Abbott, E.C. Martens, H.J. Gilbert, Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism, Nature. 517 (2015) 165–169. doi:10.1038/nature13995.

[15] J.N. D'Elia, A.A. Salyers, Contribution of a neopullulanase, a pullulanase, and an α-glucosidase to growth of Bacteroides thetaiotaomicron on starch., J Bacteriol. 178 (1996) 7173–7179.

[16] M. Okuyama, K. Matsunaga, K. Watanabe, K. Yamashita, T. Tagami, A. Kikuchi, M. Ma, P. Klahan, H. Mori, M. Yao, A. Kimura, Efficient synthesis of α-galactosyl oligosaccharides using a mutant Bacteroides thetaiotaomicron retaining α-galactosidase (BtGH97b), FEBS J. 284 (2017) 766–783. doi:10.1111/febs.14018.

[17] W. Li, H. Fan, C. He, X. Zhang, X. Wang, J. Yuan, Z. Fang, W. Fang, Y. Xiao, PspAG97A: A halophilic α-glucoside hydrolase with wide substrate specificity from glycoside hydrolase family 97, J Microbiol Biotechn. 26 (2016) 1933–1942. doi:10.4014/jmb.1606.06047.

[18] C. He, J. Li, W. Li, Y. Xue, Z. Fang, W. Fang, X. Zhang, X. Wang, Y. Xiao, Structures of PspAG97A α-glucoside hydrolase reveal a novel mechanism for chloride induced activation, J Struct Biol. 196 (2016) 426–436. doi:10.1016/j.jsb.2016.09.009.

[19] S. Yu, K. Bojsen, B. Svensson, J. Marcussen, α-1,4-Glucan lyases producing 1,5-anhydro-d-fructose from starch and glycogen have sequence similarity to α-glucosidases,

BBA-Protein Struct M. 1433 (1999) 1–15. doi:10.1016/S0167-4838(99)00152-1.

[20] M. Okuyama, W. Saburi, H. Mori, A. Kimura, α-Glucosidases and α-1,4-glucan lyases: structures, functions, and physiological actions, Cellular and Molecular Life Sciences. 73 (2016) 2727–2751. doi:10.1007/s00018-016-2247-5.

[21] J. Larsbrink, A. Izumi, G.R. Hemsworth, G.J. Davies, H. Brumer, Structural enzymology of Cellvibrio japonicus Agd31B protein reveals α-transglucosylase activity in glycoside hydrolase family 31, J Biol Chem. 287 (2012) 43288–43299. doi:10.1074/jbc.M112.416511.

[22] T. Miyazaki, Y. Ishizaki, M. Ichikawa, A. Nishikawa, T. Tonozuka, Structural and biochemical characterization of novel bacterial α-galactosidases belonging to glycoside hydrolase family 31, Biochem J. 469 (2015) 145–158. doi:10.1042/bj20150261.

[23] G. Speciale, Y. Jin, G.J. Davies, S.J. Williams, E.D. Goddard-Borger, YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids., Nat Chem Biol. 12 (2016) 215–217. doi:10.1038/nchembio.2023.

[24] T. Tagami, E. Miyano, J. Sadahiro, M. Okuyama, T. Iwasaki, A. Kimura, Two novel glycoside hydrolases responsible for the catabolism of cyclobis-(1→6)-α-nigerosyl, J Biol Chem. 291 (2016) 16438–16447. doi:10.1074/jbc.M116.727305.

[25] S.H. Light, L.A. Cahoon, K. V. Mahasenan, M. Lee, B. Boggess, A.S. Halavaty, S. Mobashery, N.E. Freitag, W.F. Anderson, Transferase versus hydrolase: The role of conformational flexibility in reaction specificity, Structure. 25 (2017) 295–304. doi:10.1016/j.str.2016.12.007.

[26] Y. Gozu, Y. Ishizaki, Y. Hosoyama, T. Miyazaki, A. Nishikawa, T. Tonozuka, A glycoside hydrolase family 31 dextranase with high transglucosylation activity from Flavobacterium johnsoniae, Biosci Biotechnol Biochem. 80 (2016) 1562–1567. doi:10.1080/09168451.2016.1182852.

[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, Skylign: 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. Altermann, 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. Kalemanov, M. Kluge, A. Meier, H. Nasir, U. Neumaier, V. Prade, J. Reeb, A. Sorokoumov, I. Troshani, S. Vorberg, S. Waldraff, 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.

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

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

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

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

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

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

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

649 doi:10.1371/journal.pbio.1001221.
650 [43] S. Vidal, T. Doco, P. Williams, P. Pellerin, W.S. York, M.A. O'Neill, J. Glushka, A.G.
651 Darvill, P. Albersheim, Structural characterization of the pectic polysaccharide
652 rhamnogalacturonan II: evidence for the backbone location of the aceric acid-containing
653 oligoglycosyl side chain, Carbohydr Res Res. 326 (2000) 277–294. doi:10.1016/S0008-
654 6215(00)00036-7.
655 [44] J.N. Glushka, M. Terrell, W.S. York, M.A. O'Neill, A. Gucwa, A.G. Darvill, P.
656 Albersheim, J.H. Prestegard, Primary structure of the 2-O-methyl-α-l-fucose-containing
657 side chain of the pectic polysaccharide, rhamnogalacturonan II, Carbohydr Res Res. 338
658 (2003) 341–352. doi:10.1016/S0008-6215(02)00461-5.
659 [45] H. Ichinose, Z. Fujimoto, M. Honda, K. Harazono, Y. Nishimoto, A. Uzura, S. Kaneko,
660 A β-L-Arabinopyranosidase from Streptomyces avermitilis is a novel member of
661 glycoside hydrolase family 27., J Biol Chem. 284 (2009) 25097–25106.
662 doi:10.1074/jbc.M109.022723.
663 [46] S. Lansky, R. Salama, H. V. Solomon, H. Feinberg, H. Belrhali, Y. Shoham, G. Shoham,
664 Structure–specificity relationships in Abp, a GH27 β-L-arabinopyranosidase from
665 Geobacillus stearothermophilus T6, Acta Crystallogr D Biol Crystallogr. 70 (2014)
666 2994–3012. doi:10.1107/S139900471401863X.
667 [47] S.-P. Nie, C. Wang, S.W. Cui, Q. Wang, M.-Y. Xie, G.O. Phillips, The core
668 carbohydrate structure of Acacia seyal var. seyal (Gum arabic), Food Hydrocolloids. 32
669 (2013) 221–227. doi:10.1016/j.foodhyd.2012.12.027.
670 [48] S.-P. Nie, C. Wang, S.W. Cui, Q. Wang, M.-Y. Xie, G.O. Phillips, A further amendment
671 to the classical core structure of gum arabic (Acacia senegal), Food Hydrocolloids. 31
672 (2013) 42–48. doi:10.1016/j.foodhyd.2012.09.014.
673 [49] L. Lopez-Torrez, M. Nigen, P. Williams, T. Doco, C. Sanchez, Acacia senegal vs.
674 Acacia seyal gums - Part 1: Composition andstructure 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 mg·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-Ara*p* (1→3) L-Ara*f* (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.

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

Fig. 5. Substrate-binding pockets of BT3661 (A) and BT1871 (B), and alignments of the sequences forming the +1 subsite (C). Secondary structure elements and residues numbers on the top and bottom of the alignment were those of BT1871 and BT3661, respectively. Similarity colors are based on the BLOSUM62 scoring matrix. The residues relevant to the aromatic clamp 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	*P*1
Unit cell parameters *a*, *b*, *c* (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]
R_{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
R_{work}/R_{free} (%)	20.7/25.3
No. of atoms	
Macromolecules	20,162
Ligand/ion	4
Water	1,257
B-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.

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

**$R_{work} = \Sigma_{hkl} ||F_{obs}| - |F_{calc}|| / \Sigma_{hkl} |Fobs|$, 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 (μmol/min/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} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (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

[a] These values were obtained from hydrolytic rates at low substrate concentrations of 1−2.5 mM.

633

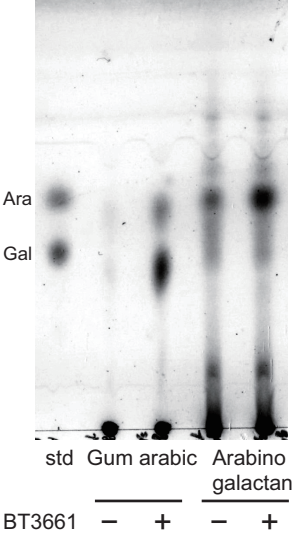


Fig. 1. Kikuchi et al

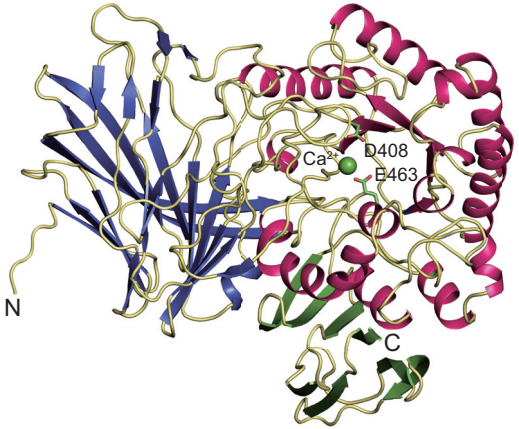


Fig. 2. Kikuchi et al

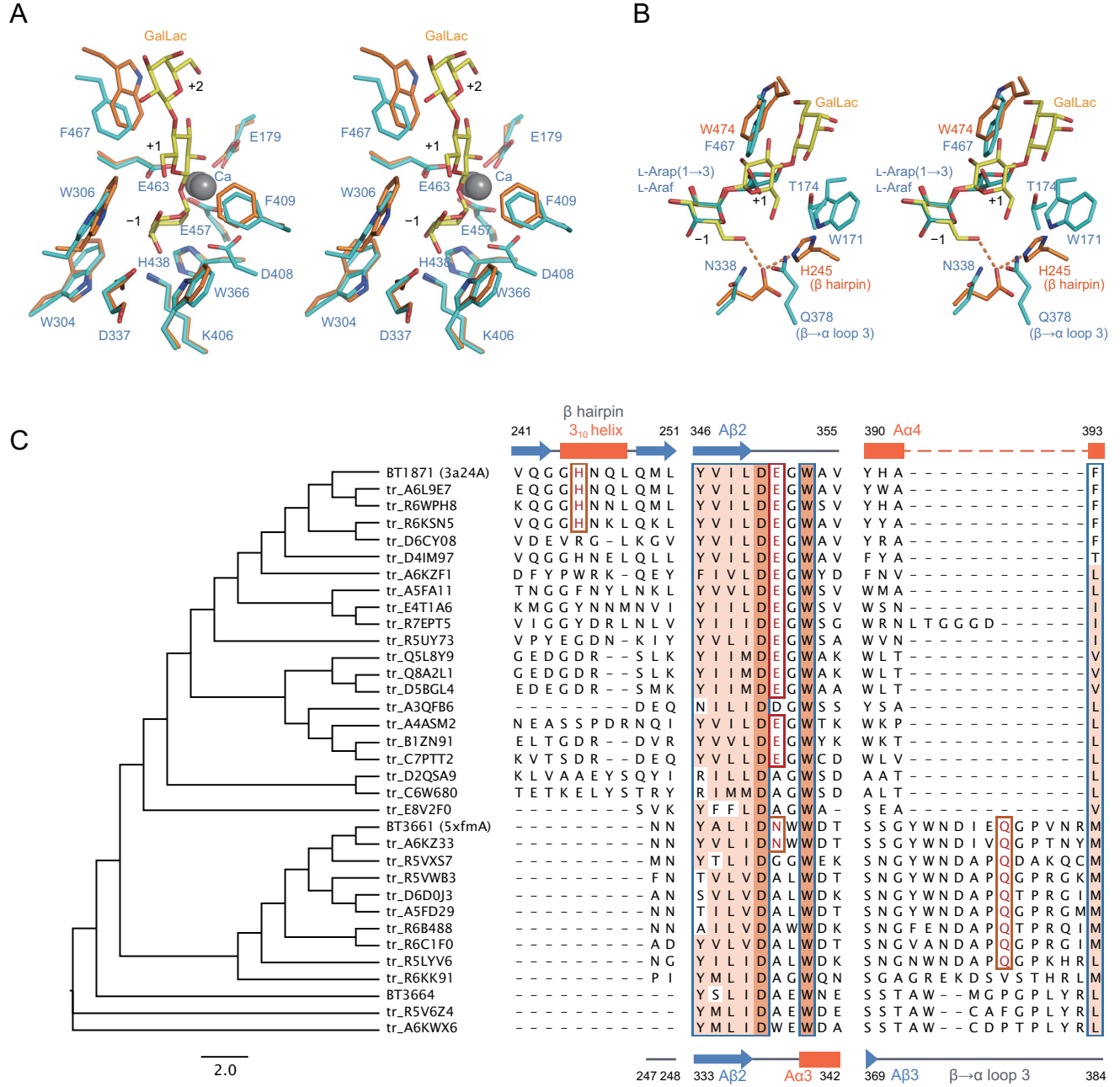


Fig. 3. Kikuchi et al

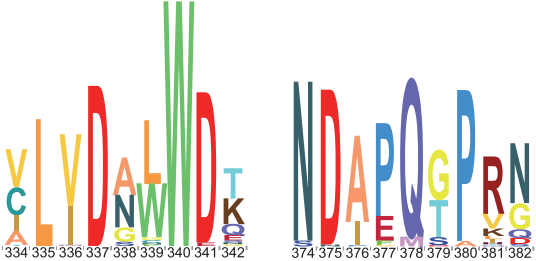
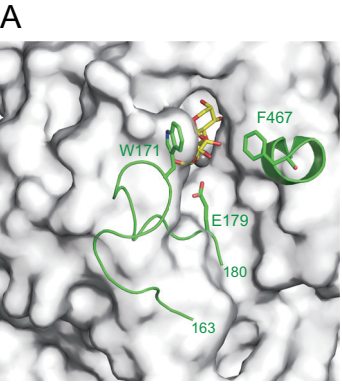
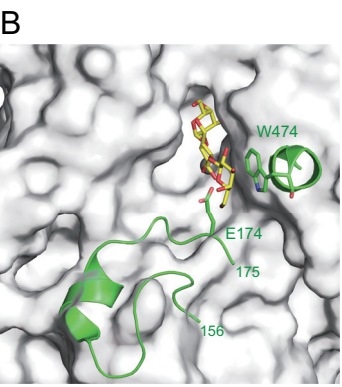


Fig. 4 Kikuchi et al





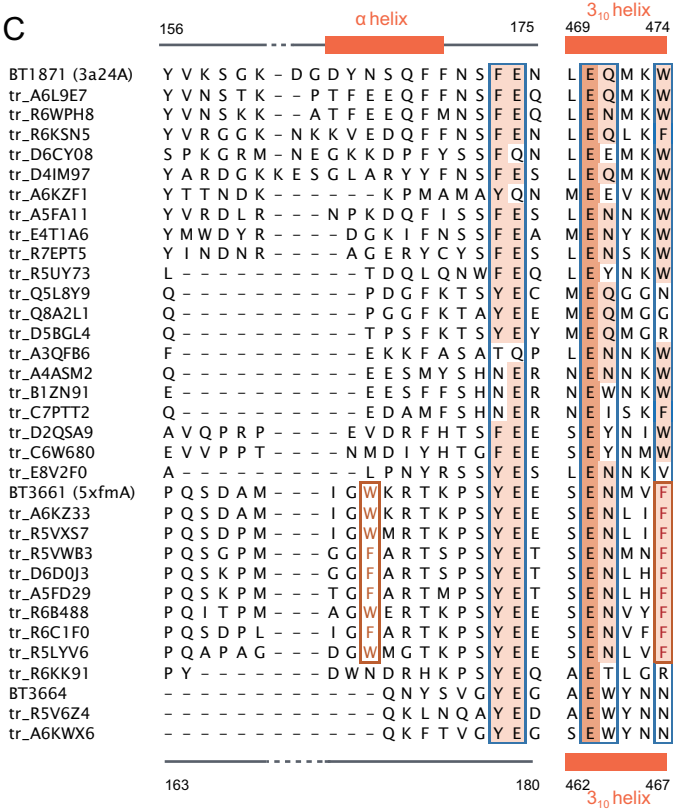


Fig. 5. Kikuchi et al

Biochimie

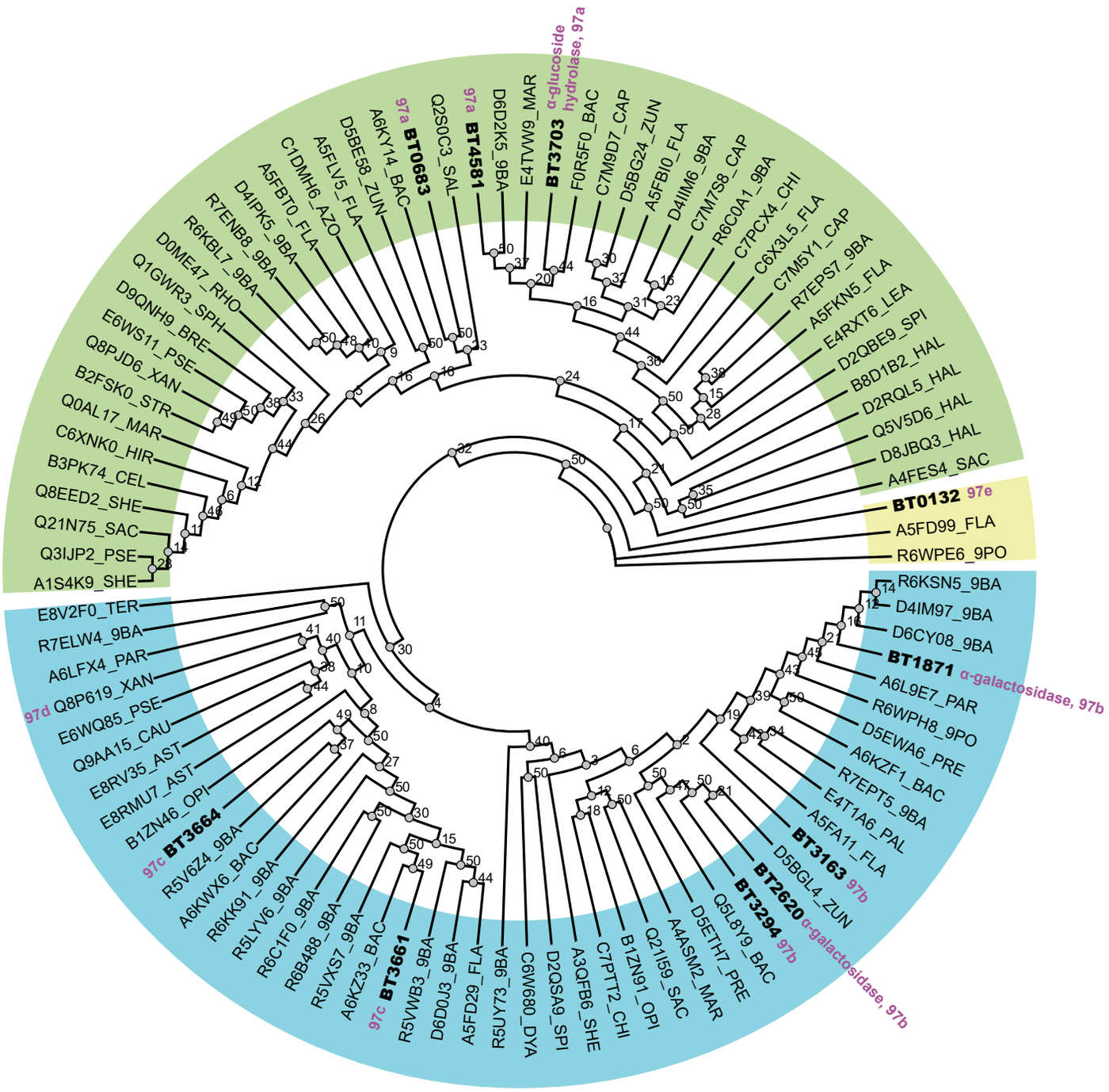
Supplementary materials for

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

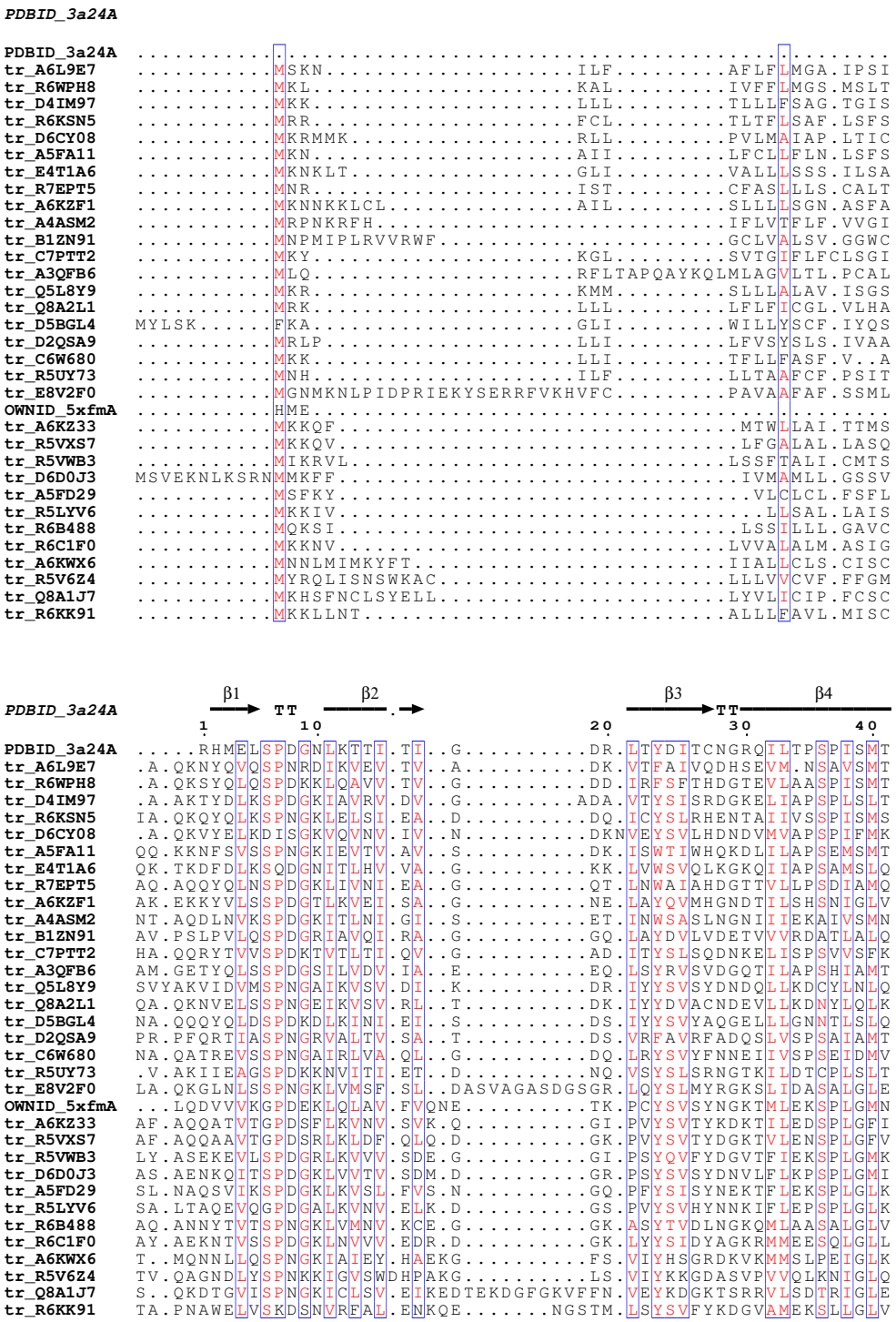
Asako Kikuchi[a,†], Masayuki Okuyama[a,*,†], Koji Kato[b], Shohei Osaki[a], Min Ma[a], Yuya Kumagai[a], Kana Matsunaga[a], Patcharapa Klahan[a], Takayoshi Tagami[a], Min Yao[b], Atsuo Kimura[a,*]

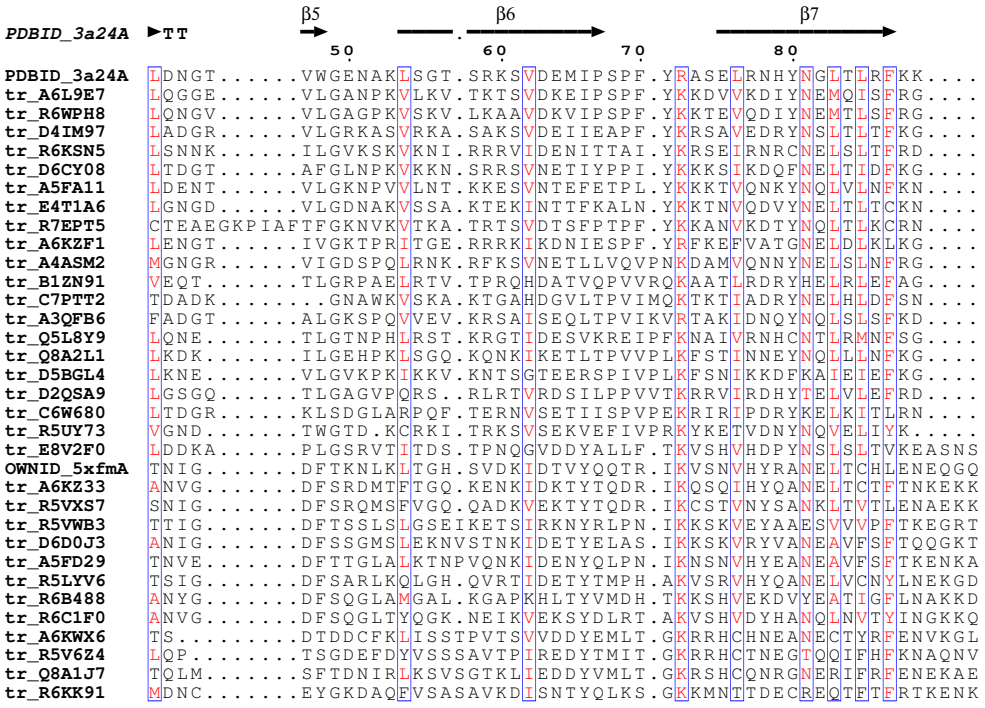
[a] Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan. Kita-9 Nishi-9, Kita-ku, Sapporo 060-8589, Japan.
[b] Faculty of Advance Life Science, Hokkaido University, Sapporo, Japan. Kita-10 Nishi-8, Kita-ku, Sapporo 060-0810, Japan.

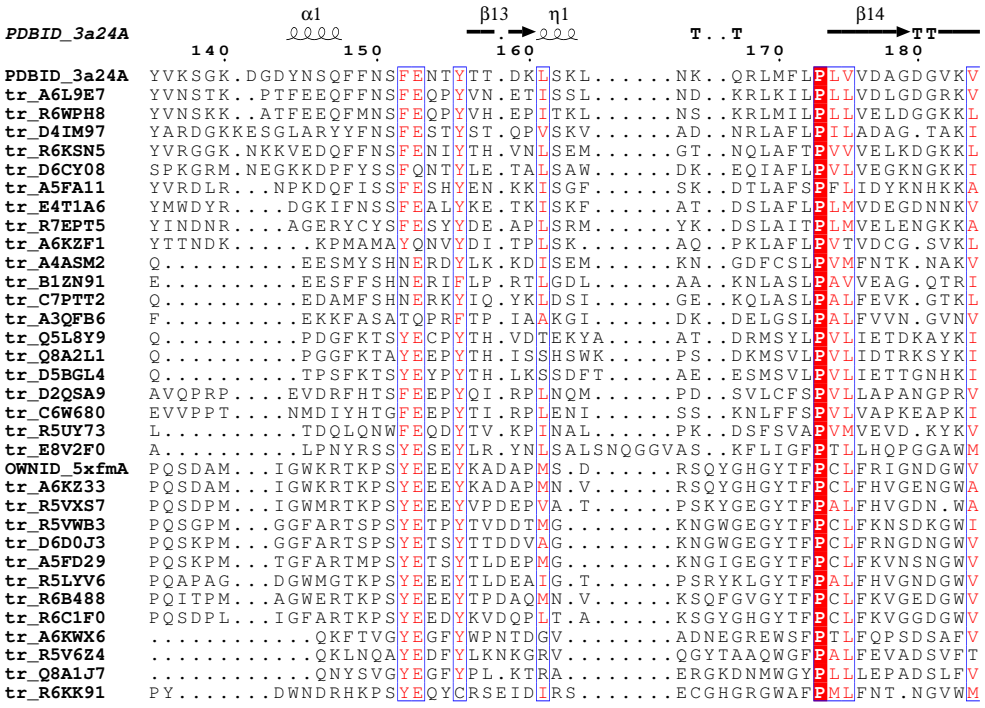
*Corresponding authors: M Okuyama, okuyama@abs.agr.hokudai.ac.jp; A Kimura, kimura@abs.agr.hokudai.ac.jp.

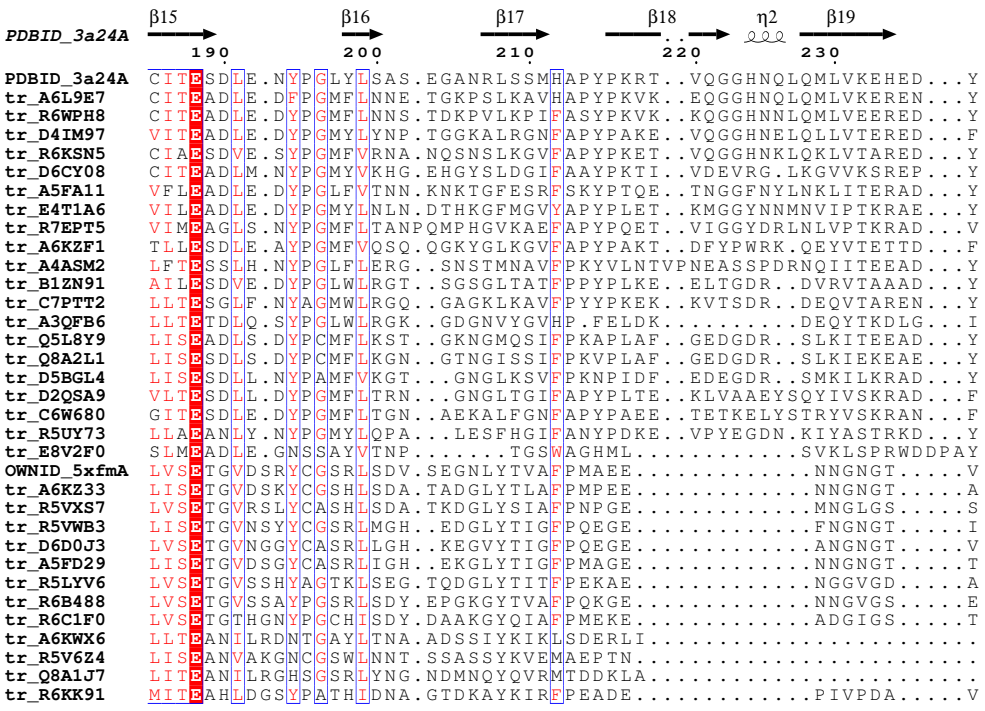


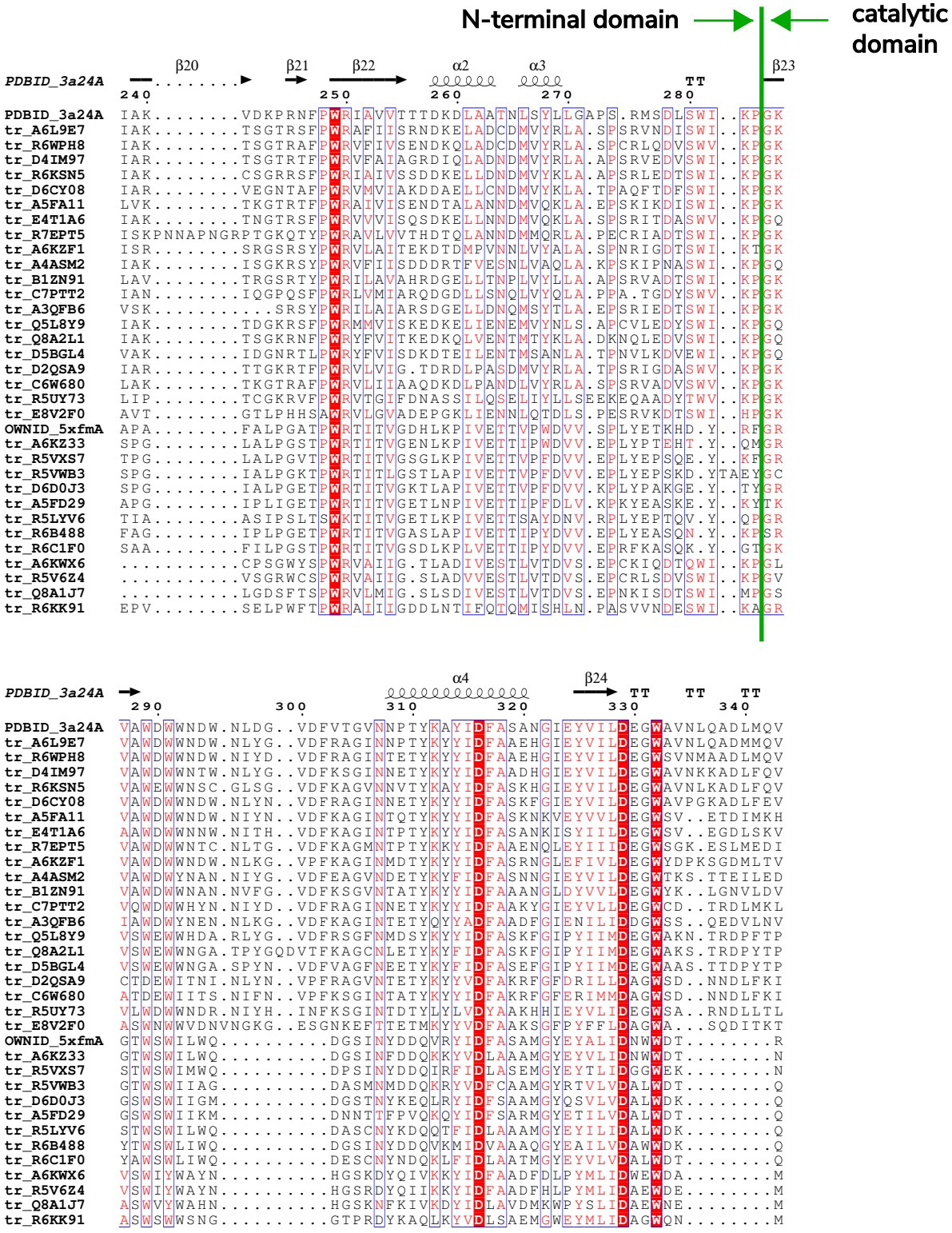
Supplementary Fig. 1. Phylogenetic tree of GH97 members. Amino acid sequences of the catalytic domain (PF10566) were aligned by Clustal omega. A maximum likelihood phylogeny was estimated from the alignment obtained using PhyML-SMS. Amino acid sequences having Glu at β-strand 5, Asp at β-strand 4, and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

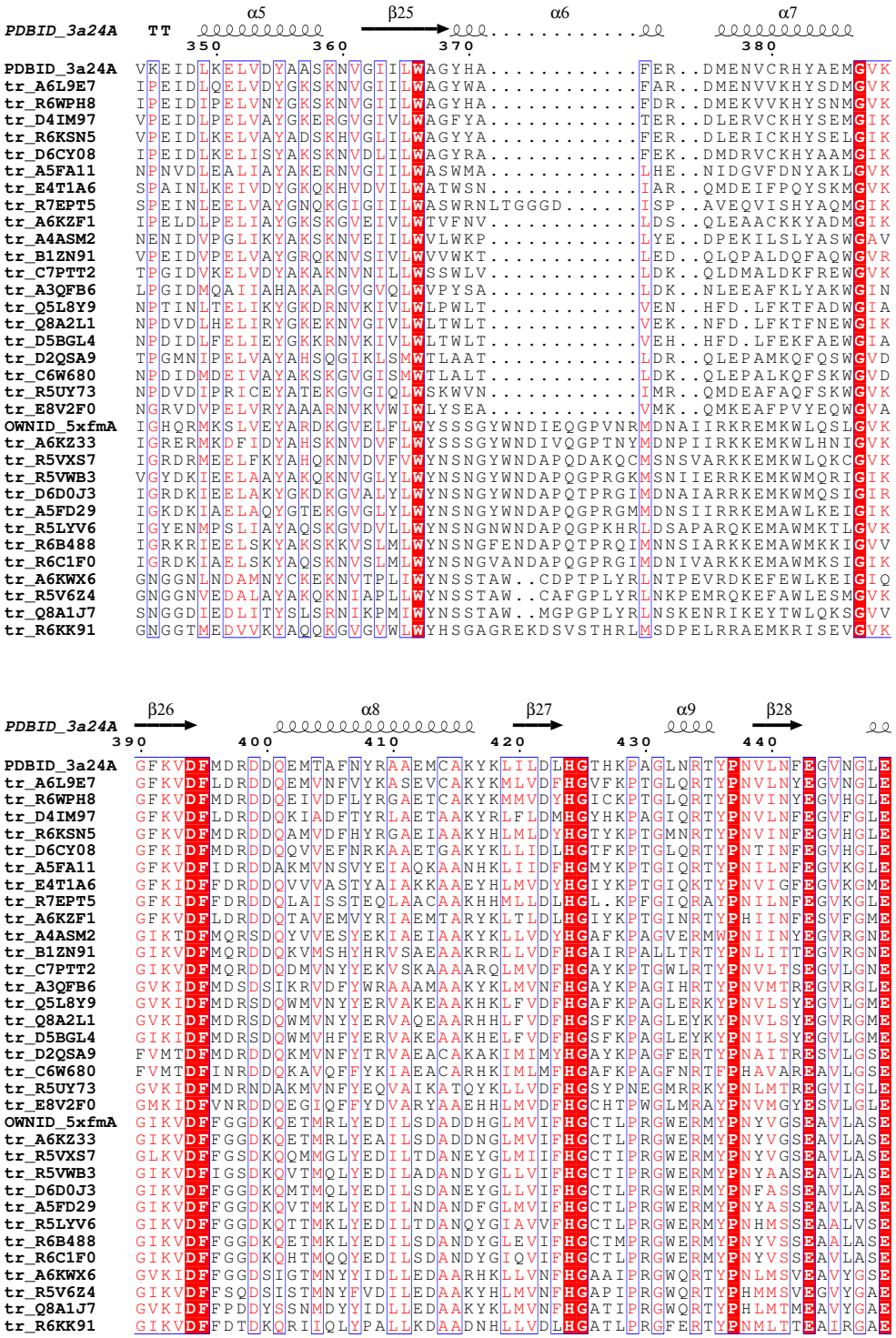

PDBID_3a24A
PDBID_3a24A
tr_A6L9E7
tr_R6WPH8
tr_D4IM97
tr_R6KSN5
tr_D6CY08
tr_A5FA11
tr_E4T1A6
tr_R7EPT5
tr_A6KZF1
tr_A4ASM2
tr_B1ZN91
tr_C7PTT2
tr_A3QFB6
tr_Q5L8Y9
tr_Q8A2L1
tr_D5BGL4
tr_D2QSA9
tr_C6W680
tr_R5UY73
tr_E8V2F0
OWNID_5xfmA
tr_A6KZ33
tr_R5VXS7
tr_R5VWB3
tr_D6D0J3
tr_A5FD29
tr_R5LYV6
tr_R6B488
tr_R6C1F0
tr_A6KWX6
tr_R5V6Z4
tr_Q8A1J7
tr_R6KK91
β1
β2
β3
β4
PDBID_3a24A
1
10
20
30
40
PDBID_3a24A
tr_A6L9E7
tr_R6WPH8
tr_D4IM97
tr_R6KSN5
tr_D6CY08
tr_A5FA11
tr_E4T1A6
tr_R7EPT5
tr_A6KZF1
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

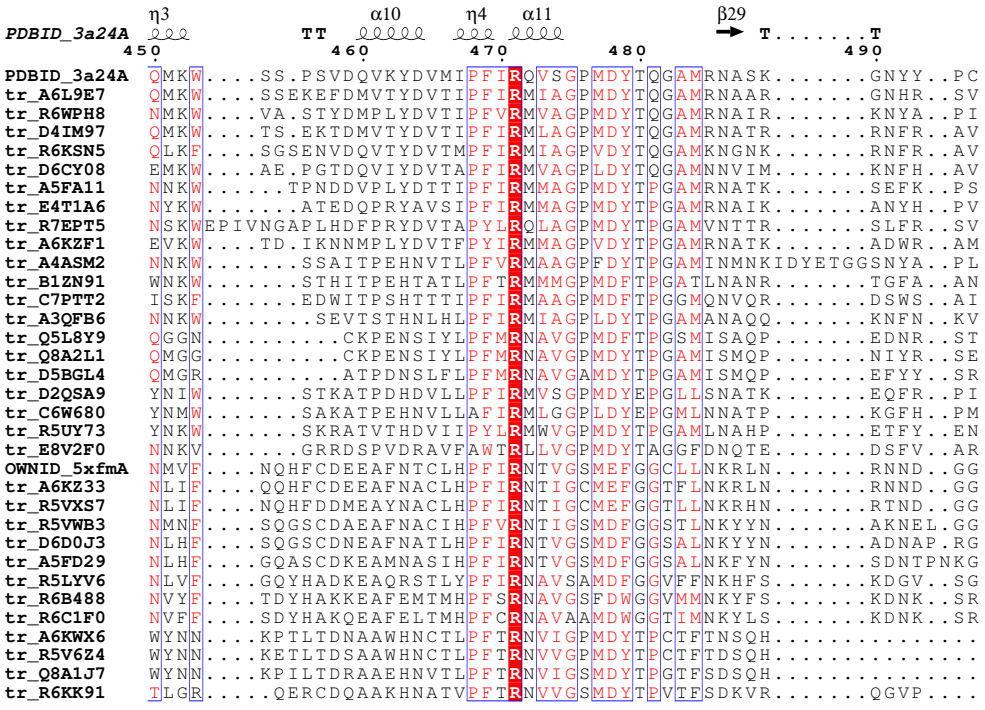

β5
β6
β7
PDBID_3a24A
50
60
70
80
PDBID_3a24A LDNGT.....VWGENAKLSGT.SRKSVDEMIPSPF.YRASELRNHYNGLTLRFKK....
tr_A6L9E7 LQGGE.....VLGANPKVLKV.TKTSVDKEIPSPF.YKKDVVKDIYNEMQISFRG....
tr_R6WPH8 LQNGV.....VLGAGPKVSKV.LKAAVDKVIPSPF.YKKTEVQDIYNEMTLSFRG....
tr_D4IM97 LADGR.....VLGRKASVRKA.SAKSVDEIIEAPF.YKRSAVEDRYNSLTLTFKG....
tr_R6KSN5 LSNNK.....ILGVKSKVKNI.RRRVIDENITTAI.YKRSEIRNRCNELSLTFRD....
tr_D6CY08 LTDGT.....AFGLNPKVKKN.SRRSVNETIYPPI.YKKKSIKDQFNELTIDFKG....
tr_A5FA11 LDENT.....VLGKNPVVLNT.KKESVNTEFETPL.YKKKTVQNKYNQLVLNFKN....
tr_E4T1A6 LGNGD.....VLGDNAKVSSA.KTEKINTTFKALN.YKKTNVQDVYNELTLTCKN....
tr_R7EPT5 CTEAEGKPIAFTFGKNVKVTKA.TRTSVDTSFPTPF.YKKANVKDTYNQLTLKCRN....
tr_A6KZF1 LENGT.....IVGKTPRITGE.RRRKIKDNIESPF.YRFKEFVATGNELDLKLKG....
tr_A4ASM2 MGNGR.....VIGDSPQLRNK.RFKSVNETLLVQVPNKDAMVQNNYNELSLNFRG....
tr_B1ZN91 VEQT.....TLGRPAELRTV.TPRQHDATVQPVVRQKAATLRDRYHELRLEFAG....
tr_C7PTT2 TDADK.....GNAWKVSKA.KTGAHDGVLTPVIMQKTKTIADRYNELHLDFSN....
tr_A3QFB6 FADGT.....ALGKSPQVVEV.KRSAISEQLTPVIKVRTAKIDNQYNQLSLSFKD....
tr_Q5L8Y9 LQNE.....TLGTNPHLRST.KRGTIDESVKREIPFKNAIVRNHCNTLRMNFSG....
tr_Q8A2L1 LKDK.....ILGEHPKLSGQ.KQNKIKETLTPVVPLKFSTINNEYNQLLLNFKG....
tr_D5BGL4 LKNE.....VLGVKPKIKKV.KNTSGTEERSPIVPLKFSNIKKDFKAIEIEFKG....
tr_D2QSA9 LGSGQ.....TLGAGVPQRS..RLRTVRDSILPPVVTKRRVIRDHYTELVLEFRD....
tr_C6W680 LTDGR.....KLSDGLARPQF.TERNVSETIISPVPEKRIRIPDRYKELKITLRN....
tr_R5UY73 VGND.....TWGTD.KCRKI.TRKSVSEKVEFIVPRKYKETVDNYNQVELIYK.....
tr_E8V2F0 LDDKA.....PLGSRVTITDS.TPNQGVDDYALLF.TKVSHVHDPYNSLSLTVKEASNS
OWNID_5xfmA TNIG.....DFTKNLKLTGH.SVDKIDTVYQQTR.IKVSNVHYRANELTCHLENEQGQ
tr_A6KZ33 ANVG.....DFSRDMTFTGQ.KENKIDKTYTQDR.IKQSQIHYQANELTCTFTNKEKK
tr_R5VXS7 SNIG.....DFSRQMSFVGQ.QADKVEKTYTQDR.IKCSTVNYSANKLTVTLENAEKK
tr_R5VWB3 TTIG.....DFTSSLSLGSEIKETSIRKNYRLPN.IKKSKVEYAAESVVVPFTKEGRT
tr_D6D0J3 ANIG.....DFSSGMSLEKNVSTNKIDETYELAS.IKKSKVRYVANEAVFSFTQQGKT
tr_A5FD29 TNVE.....DFTTGLALKTNPVQNKIDENYQLPN.IKNSNVHYEANEAVFSFTKENKA
tr_R5LYV6 TSIG.....DFSARLKQLGH.QVRTIDETYTMPH.AKVSRVHYQANELVCNYLNEKGD
tr_R6B488 ANYG.....DFSQGLAMGAL.KGAPKHLTYVMDH.TKKSHVEKDVYEATIGFLNAKKD
tr_R6C1F0 ANVG.....DFSQGLTYQGK.NEIKVEKSYDLRT.AKVSHVDYHANQLNVTYINGKKQ
tr_A6KWX6 TS.....DTDDCFKLISSTPVTSVVDDYEMLT.GKRRHCHNEANECTYRFENVKGL
tr_R5V6Z4 LQP.....TSGDEFDYVSSSAVTPIREDYTMIT.GKRRHCTNEGTQQIFHFKNAQNV
tr_Q8A1J7 TQLM.....SFTDNIRLKSVSGTKLIEDDYVMLT.GKRSHCQNRGNERIFRFENEKAE
tr_R6KK91 MDNC.....EYGKDAQFVSASAVKDISNTYQLKS.GKKMNTTDECREQTFTFRTKENK
β8
β9
β10
β11
β12
PDBID_3a24A
90
100
110
120
130
PDBID_3a24A .....DWNVEFRAYNDGIAYRFVNQG....KK..PFR...VVTEVSDYCFPSDMTASVP
tr_A6L9E7 .....DYGLVFRVYNDGVAYRFTTKR....KE..PIV...IADEEAVYNFPSDYKTFTA
tr_R6WPH8 .....NYGLVFRMYNDGLAYRFTTKM....KN..DIV...VVDEEADYTFSSDHMAFAP
tr_D4IM97 .....DYGIDFRAYDDGAAYRFRTSM....KD..EIT...VTDELVDIRFPEDFTTLVP
tr_R6KSN5 .....GFQLEFRAYDEGVAYRFVVTE....NK..PFN...VIAEEATFNFDDDYMVYIP
tr_D6CY08 .....GYSLVFRAYEDGAAYRFVSEL....KK..PFM...VESEQASFCFPNDPKVFVA
tr_A5FA11 .....DFSIEFRVFDDGAAYRFITKK....KK..NIT...VKWENVSLNFDQDYNTLMP
tr_E4T1A6 .....DFGVKFRVYNDGVAYRFFTKK....KG..EIV...IKNEEANFNFTQDYKAFIP
tr_R7EPT5 .....GYSIQFRAYDDGAAYRFVSEL....RK..PFC...VKDEIAEFNFDRDYHAFVP
tr_A6KZF1 .....GFGIIFRAYNEGVAYRFYTTQ....SS..DII...IKEEQAEFNFKEDYTAYLP
tr_A4ASM2 .....EYQLLFRAFDDGIAYQFIDKS....SR..TKK...VVNEQMSLTFPEGTSSFFP
tr_B1ZN91 .....NYAVVFRAYDEGVAYRWETQLA...QS..EVK...VFAEEAAFNFADNQTVWYP
tr_C7PTT2 .....GLALEWRAYDNGIAWHWLTTS....KK..PYK...VLDEQAGFSLDKEGKAWYP
tr_A3QFB6 .....NYQVDFRAFDEGVAYRFSAQQ....SG..KSA...LTSEQAEFNFAKGAFAYFP
tr_Q5L8Y9 .....NYAVEFRVFDNGIAYRFVTDK....KG..DNI...VMGEDFAINFPTNYKAHLS
tr_Q8A2L1 .....NYSVEFRAFNDGIAYRFITRQ....KG..EIE...VLGEDLSLQFPTDYLLHLQ
tr_D5BGL4 .....NYSIEFKAYNEGIAYRFKLGK....KG..SLE...VIQENFDIVFPAEYTLHLQ
tr_D2QSA9 .....QHALEFRVYDDGVAYRFRTAF....TD..SIT...VRTETARFAFADNPTVYYP
tr_C6W680 .....AFTVIFRVYDDGVAYRIATSL....PG..EIT...VRSETAAFRFPANHPVYFP
tr_R5UY73 .....DYKIEFRIYNDGVAYRFVSTA....NN..KQP...VKKESVSFRFGQDHTSYTL
tr_E8V2F0 GGRNGRSMVVEARAYDTGIAFRYVLPS....QA..SITQLHLRGEQTEFRFSQDDTAWVL
OWNID_5xfmA .....KLGVIFRVSDNDVAFRYTLPHQG..GK.ASVT...VKEEQTGFRFPEQTTTFLC
tr_A6KZ33 .....NINIIFRVSNNDIAFRYEMPKYG..DT.GSIV...IEKETTGFDFPSFTTTFLC
tr_R5VXS7 .....KLDIVFQLSNNDIAFRYEMPQYG..ET.ACMV...IEKEATGFDFPSTTTTFLS
tr_R5VWB3 .....AFDVEFRVSNNDVAFRYTVYPKK..DI.LCCV...IQQETTGFVMPEGTTTFLC
tr_D6D0J3 .....IYDVIFRISNNDVAFKYKMYPQG..ET.LSCV...VKQEVTGFAFPDGTTTFLC
tr_A5FD29 .....AIDIIFRVSNNNAAFKYKVYPQK..ES.RSCV...VEGEASGFLLPEGTTTFLC
tr_R5LYV6 .....TLQVIFRVSNNDIAQAYQISSPK..HT..HCI...IEKETTGFDFPSYTTTFIT
tr_R6B488 .....SMTLHLHVSDNDVAYKYEMSRPKKDNP.KSVI...IYKEVSGFNFPEKTTTFLC
tr_R6C1F0 .....PMTVTFNVSNNDVAFRYTFDQL...KA.QHLI...VNEEKTAFNLPKVTTTYLC
tr_A6KWX6 .....RVDLIFRVYNDGIAFRYHLPK....TK.EEIV...VLDEYTAFAFTPGIKRWT.
tr_R5V6Z4 .....SLDLELRVYDDGIAFRYVFPQ....SA.KEIC...VTDETTAFVFQEGITRWT.
tr_Q8A1J7 .....TLDIVFRAYDDGIVFRYSLPS....VQ.DKDS...LTSEHTAYYIPEGTKRWI.
tr_R6KK91 .....QFNLIVRVYNDGVAFRYELPG....EDGQMHT...IQAEHTEFAVPVNGKAWIH

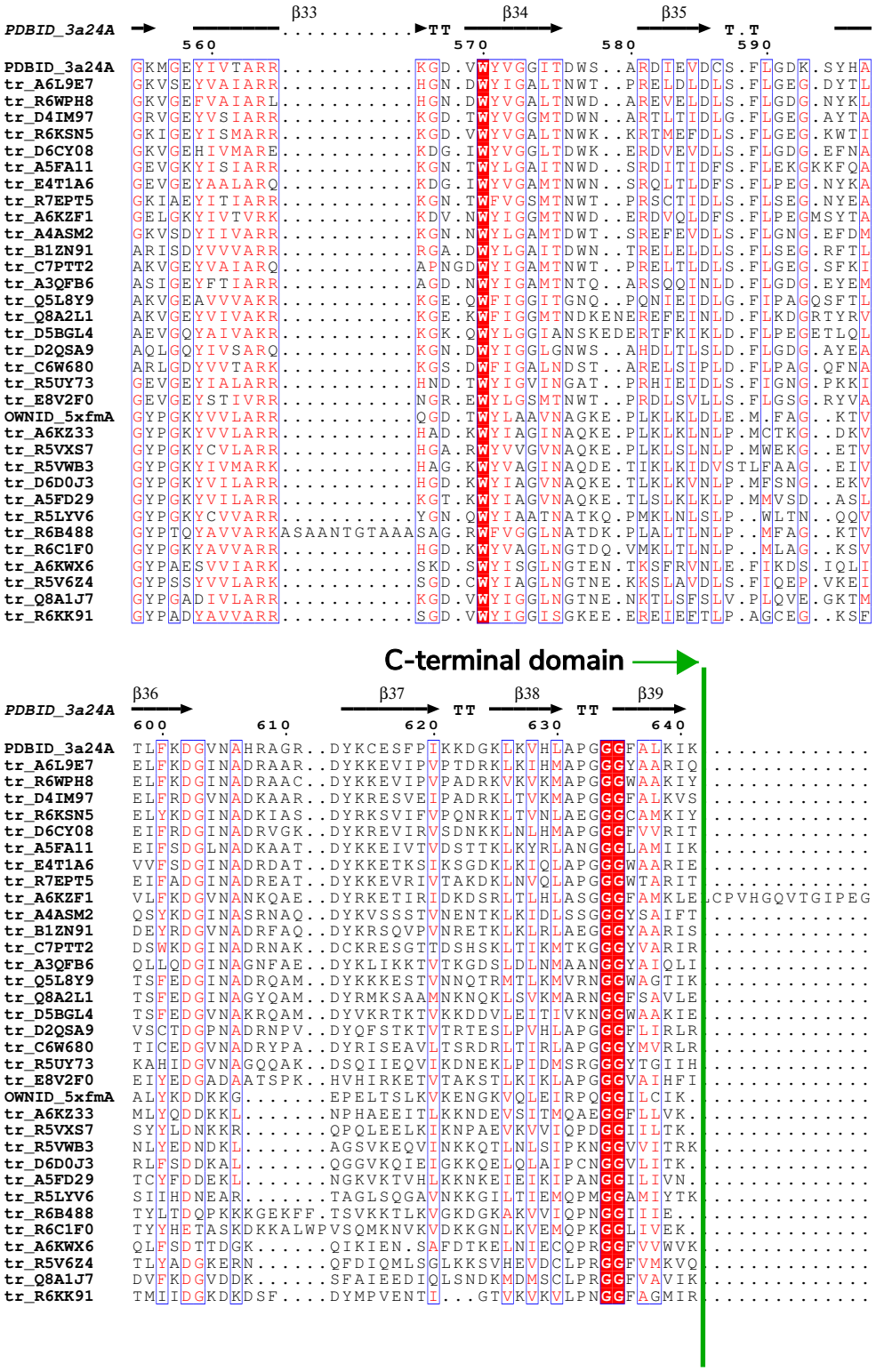

PDBID_3a24A
α1
β13
η1
β14
140
150
160
170
180
PDBID_3a24A YVKSGK.DGDYNSQFFNSFENTYTT.DKLSKL.....NK..QRLMFLPLVVDAGDGVKV
tr_A6L9E7 YVNSTK..PTFEEQFFNSFEQPYVN.ETISSL.....ND..KRLKILPLLVDLGDGRKV
tr_R6WPH8 YVNSKK..ATFEEQFMNSFEQPYVH.EPITKL.....NS..KRLMILPLLVELDGGKKL
tr_D4IM97 YARDGKKESGLARYYFNSFESTYST.QPVSKV.....AD..NRLAFLPILADAG.TAKI
tr_R6KSN5 YVRGGK.NKKVEDQFFNSFENITYTH.VNLSEM.....GT..NQLAFTPVVVELKDGKKL
tr_D6CY08 SPKGRM.NEGKKDPFYSSFQNTYLE.TALSAW.....DK..EQIAFLPVLVEGKNGKKI
tr_A5FA11 YVRDLR...NPKDQFISSFESHYEN.KKISGF.....SK..DTLAFSPFLIDYKNHKKA
tr_E4T1A6 YMWDYR....DGKIFNSSFEALYKE.TKISKF.....AT..DSLAFLPLMVDEGDNNKV
tr_R7EPT5 YINDNR....AGERYCYSFESYYDE.APLSRM.....YK..DSLAITPLMVELENGKKA
tr_A6KZF1 YTTNDK.....KPMAMAYQNVYDI.TPLSK.....AQ..PKLAFLPVTVDCG.SVKL
tr_A4ASM2 Q.....EESMYSHNERDYLK.KDISEM.....KN..GDFCSLPVMFNTK.NAKV
tr_B1ZN91 E.....EESFFSHNERIFLP.RTLGDL.....AA..KNLASLPAVVEAG.QTRI
tr_C7PTT2 Q.....EDAMFSHNERKYIQ.YKLDSI.....GE..KQLASLPALFEVK.GTKL
tr_A3QFB6 F.....EKKFASATQPRFTP.IAAKGI.....DK..DELGSLPALFVVN.GVNV
tr_Q5L8Y9 Q.....PDGFKTSYECPYTH.VDTEKYA.....AT..DRMSYLPVLIETDKAYKI
tr_Q8A2L1 Q.....PGGFKTAYEEPYTH.ISSHSWK.....PS..DKMSVLPVLIDTRKSYKI
tr_D5BGL4 Q.....TPSFKTSYEYPYTH.LKSSDFT.....AE..ESMSVLPVLIETTGNHKI
tr_D2QSA9 AVQPRP....EVDRFHTSFEEPYQI.RPLNQM.....PD..SVLCFSPVLLAPANGPRV
tr_C6W680 EVVPPT....NMDIYHTGFEEPYTI.RPLENI.....SS..KNLFFSPVLVAPKEAPKI
tr_R5UY73 L.....TDQLQNWFEQDYTV.KPINAL.....PK..DSFSVAPVMVEVD.KYKV
tr_E8V2F0 A.....LPNYRSSYESEYLR.YNLSALSNQGGVAS..KFLIGFPTLLHQPGGAWM
OWNID_5xfmA PQSDAM...IGWKRTKPSYEEEYKADAPMS.D.....RSQYGHGYTFPCLFRIGNDGWV
tr_A6KZ33 PQSDAM...IGWKRTKPSYEEEYKADAPMN.V.....RSQYGHGYTFPCLFHVGENGWA
tr_R5VXS7 PQSDPM...IGWMRTKPSYEEEYVPDEPVA.T.....PSKYGEGYTFPALFHVGDN.WA
tr_R5VWB3 PQSGPM...GGFARTSPSYETPYTVDDTMG.....KNGWGEGYTFPCLFKNSDKGWI
tr_D6D0J3 PQSKPM...GGFARTSPSYETSYTTDDVAG.....KNGWGEGYTFPCLFRNGDNGWV
tr_A5FD29 PQSKPM...TGFARTMPSYETSYTLDEPMG.....KNGIGEGYTFPCLFKVNSNGWV
tr_R5LYV6 PQAPAG...DGWMGTKPSYEEEYTLDEAIG.T.....PSRYKLGYTFPALFHVGNDGWV
tr_R6B488 PQITPM...AGWERTKPSYEEEYTPDAQMN.V.....KSQFGVGYTFPCLFKVGEDGWV
tr_R6C1F0 PQSDPL...IGFARTKPSYEEDYKVDQPLT.A.....KSGYGHGYTFPCLFKVGGDGWV
tr_A6KWX6 .....QKFTVGYEGFYWPNTDGV.....ADNEGREWSPPTLFQPSDSAFV
tr_R5V6Z4 .....QKLNQAYEDFYLKNKGRV.....QGYTAAQWGFPALFEVADSVFT
tr_Q8A1J7 .....QNYSVGYEGFYPL.KTRA.....ERGKDNMWGYPLLLEPADSLFV
tr_R6KK91 PY.....DWNDRHKPSYEQYCRSEIDIRS.....ECGHGRGWAFPMLFNT.NGVWM


PDBID_3a24A
β15
β16
β17
β18
η2
β19
190
200
210
220
230
PDBID_3a24A CITESDLE.NYPGLYLSAS.EGANRLSSMHAPYPKRT..VQGGHNQLQMLVKEHED...Y
tr_A6L9E7 CITEADLE.DFPGMFLNNE.TGKPSLKAVHAPYPKVK..EQGGHNQLQMLVKEREN...Y
tr_R6WPH8 CITEADLE.DYPGMFLNNS.TDKPVLKPIFASYPKVK..KQGGHNNLQMLVEERED...Y
tr_D4IM97 VITEADLE.DYPGMYLYNP.TGGKALRGNFAPYPAKE..VQGGHNELQLLVTERED...F
tr_R6KSN5 CIAESDVE.SYPGMFVRNA.NQSNSLKGVFAPYPKET..VQGGHNKLQKLVTARED...Y
tr_D6CY08 CITEADLM.NYPGMYVKHG.EHGYSLDGIFAAYPKTI..VDEVRG.LKGVVKSREP...Y
tr_A5FA11 VFLEADLE.DYPGLFVTNN.KNKTGFESRFSKYPTQE..TNGGFNYLNKLITERAD...Y
tr_E4T1A6 VILEADLE.DYPGMYLNLN.DTHKGFMGVYAPYPLET..KMGGYNNMNVIPTKRAE...Y
tr_R7EPT5 VIMEAGLS.NYPGMFLTANPQMPHGVKAEFAPYPQET..VIGGYDRLNLVPTKRAD...V
tr_A6KZF1 TLLESDLE.AYPGMFVQSQ.QGKYGLKGVFAPYPAKT..DFYPWRK.QEYVTETTD...F
tr_A4ASM2 LFTESSLH.NYPGLFLERG..SNSTMNAVFPKYVLNTVPNEASSPDRNQIITEEAD...Y
tr_B1ZN91 AILESDVE.DYPGLWLRGT..SGSGLTATFPPYPLKE..ELTGDR..DVRVTAAAD...Y
tr_C7PTT2 LLTESGLF.NYAGMWLRGQ..GAGKLKAVFPYYPKEK..KVTSDR..DEQVTAREN...Y
tr_A3QFB6 LLTETDLQ.SYPGLWLRGK..GDGNVYGVHP.FELDK..........DEQYTKDLG...I
tr_Q5L8Y9 LISEADLS.DYPCMFLKST..GKNGMQSIFPKAPLAF..GEDGDR..SLKITEEAD...Y
tr_Q8A2L1 LISESDLS.DYPCMFLKGN..GTNGISSIFPKVPLAF..GEDGDR..SLKIEKEAE...Y
tr_D5BGL4 LISESDLL.NYPAMFVKGT...GNGLKSVFPKNPIDF..EDEGDR..SMKILKRAD...Y
tr_D2QSA9 VLTESDLL.DYPGMFLTRN...GNGLTGIFAPYPLTE..KLVAAEYSQYIVSKRAD...F
tr_C6W680 GITESDLE.DYPGMFLTGN..AEKALFGNFAPYPAEE..TETKELYSTRYVSKRAN...F
tr_R5UY73 LLAEANLY.NYPGMYLQPA...LESFHGIFANYPDKE..VPYEGDN.KIYASTRKD...Y
tr_E8V2F0 SLMEADLE.GNSSAYVTNP.......TGSWAGHML..........SVKLSPRWDDPAY
OWNID_5xfmA LVSETGVDSRYCGSRLSDV.SEGNLYTVAFPMAEE..........NNGNGT.....V
tr_A6KZ33 LISETGVDSKYCGSHLSDA.TADGLYTLAFPMPEE..........NNGNGT.....A
tr_R5VXS7 LVSETGVRSLYCASHLSDA.TKDGLYSIAFPNPGE..........MNGLGS.....S
tr_R5VWB3 LISETGVNSYYCGSRLMGH..EDGLYTIGFPQEGE..........FNGNGT.....I
tr_D6D0J3 LVSETGVNGGYCASRLLGH..KEGVYTIGFPQEGE..........ANGNGT.....V
tr_A5FD29 LISETGVDSGYCASRLIGH..EKGLYTIGFPMAGE..........NNGNGT.....T
tr_R5LYV6 LVSETGVSSHYAGTKLSEG.TQDGLYTITFPEKAE..........NGGVGD.....A
tr_R6B488 LVSETGVSSAYPGSRLSDY.EPGKGYTVAFPQKGE..........NNGVGS.....E
tr_R6C1F0 LVSETGTHGNYPGCHISDY.DAAKGYQIAFPMEKE..........ADGIGS.....T
tr_A6KWX6 LLTEANILRDNTGAYLTNA.ADSSIYKIKLSDERLI.....................
tr_R5V6Z4 LISEANVAKGNCGSWLNNT.SSASSYKVEMAEPTN.....................
tr_Q8A1J7 LITEANILRGHSGSRLYNG.NDMNQYQVRMTDDKLA.....................
tr_R6KK91 MITEAHLDGSYPATHIDNA.GTDKAYKIRFPEADE..........PIVPDA.....V


N-terminal domain
catalytic domain
PDBID_3a24A
β20
β21
β22
α2
α3
β23
TT
240
250
260
270
280
PDBID_3a24A IAK.....VDKPRNFPWRIAVVTTTDKDLAATNLSYLLGAPS.RMSDLSWI..KPGK
tr_A6L9E7 IAK.....TSGTRSFPWRAFIISRNDKELADCDMVYRLA.SPSRVNDISWI..KPGK
tr_R6WPH8 IAK.....TSGTRAFPWRVFIVSENDKQLADCDMVYRLA.SPCRLQDVSWV..KPGK
tr_D4IM97 IAR.....TSGTRTFPWRVFAIAGRDIQLADNDMVYRLA.APSRVEDVSWI..KPGK
tr_R6KSN5 IAK.....CSGRRSFPWRIAIVSSDDKELLDNDMVYKLA.APSRLEDTSWI..KPGK
tr_D6CY08 IAR.....VEGNTAFPWRVMVIAKDDAELLCNDMVYKLA.TPAQFTDFSWI..KPGK
tr_A5FA11 LVK.....TKGTRTFPWRAIVISENDTALANNDMVQKLA.EPSKIKDISWI..KPGK
tr_E4T1A6 IAK.....TNGTRSFPWRVVVISQSDKELLNNDMVQKLA.SPSRITDASWV..KPGQ
tr_R7EPT5 ISKPNNAPNGRPTGKQTYPWRAVLVVTHDTQLANNDMMQRLA.PECRIADTSWI..KPGK
tr_A6KZF1 ISR.....SRGSRSYPWRVLAITEKDTDMPVNNLVYALA.SPNRIGDTSWI..KTGK
tr_A4ASM2 IAK.....ISGKRSYPWRVFIISDDDRTFVESNLVAQLA.KPSKIPNASWI..KPGQ
tr_B1ZN91 LAV.....TRGSRTYPWRILAVAHRDGELLTNPLVYLLA.APSRVADTSWI..KPGK
tr_C7PTT2 IAN.....IQGPQSFPWRLVMIARQDGDLLSNQLVYQLA.PPA.TGDYSWV..KPGK
tr_A3QFB6 VSK.....SRSYPWRILAIARSDGELLDNQMSYTLA.EPSRIEDTSWI..KPGK
tr_Q5L8Y9 IAK.....TDGKRSFPWRMMVISKEDKELIENEMVYNLS.APCVLEDYSWI..KPGQ
tr_Q8A2L1 IAK.....TSGKRNFPWRYFVITKEDKQLVENTMTYKLA.DKNQLEDVSWI..KPGQ
tr_D5BGL4 VAK.....IDGNRTLPWRYFVISDKDTEILENTMSANLA.TPNVLKDVEWI..KPGQ
tr_D2QSA9 IAR.....TTGKRTFPWRVLVIG.TDRDLPASDMVYRLA.TPSRIGDASWV..KPGK
tr_C6W680 LAK.....TKGTRAFPWRVLIIAAQDKDLPANDLVYRLA.SPSRVADVSWV..KPGK
tr_R5UY73 LIP.....TCGKRVFPWRVTGIFDNASSILQSELIYLLSEEKEQAADYTWV..KPGK
tr_E8V2F0 AVT.....GTLPHHSAWRVLGVADEPGKLIENNLQTDLS.PESRVKDTSWI..HPGK
OWNID_5xfmA APA.....FALPGATPWRTITVGDHLKPIVETTVPWDVV.SPLYETKHD.Y..RPGR
tr_A6KZ33 SPG.....LALPGSTPWRTITVGENLKPIVETTIPWDVV.EPLYPTEHT.Y..QMGR
tr_R5VXS7 TPG.....LALPGVTPWRTITVGSGLKPIVETTVPFDVV.EPLYEPSQE.Y..KFGR
tr_R5VWB3 SPG.....IALPGKTPWRTITLGSTLAPIVETTVPFDVV.EPLYEPSKD.YTAEYGC
tr_D6D0J3 SPG.....IALPGETPWRTITVGKTLAPIVETTVPFDVV.KPLYPAKGE.Y..TYGR
tr_A5FD29 APG.....IPLIGETPWRTITVGETLNPIVETTIPFDLV.KPKYEASKE.Y..KYTK
tr_R5LYV6 TIA.....ASIPSLTSWKTITVGETLKPIVETTSAYDNV.RPLYEPTQV.Y..QPGR
tr_R6B488 FAG.....IPLPGETPWRTITVGASLAPIVETTIPYDVV.EPLYEASQN.Y..KPSR
tr_R6C1F0 SAA.....FILPGSTPWRTITVGSDLKPLVETTIPYDVV.EPRFKASQK.Y..GTGK
tr_A6KWX6 .....CPSGWYSPWRVAIIG.TLADIVESTLVTDVS.EPCKIQDTQWI..KPGL
tr_R5V6Z4 .....VSGRWCSPWRVAIIG.SLADVVESTLVTDVS.EPCRLSDVSWI..KPGV
tr_Q8A1J7 .....LGDSFTSPWRVLMIG.SLSDIVESTLVTDVS.EPNKISDTSWI..MPGS
tr_R6KK91 EPV.....SELPWFTPWRAIIIGDDLNTIFQTQMISHLN.PASVVNDESWI..KAGR
PDBID_3a24A
α4
β24
TT
TT
TT
290
300
310
320
330
340
PDBID_3a24A VAWDWWNDW.NLDG..VDFVTGVNNPTYKAYIDFASANGIEYVILDEGWAVNLQADLMQV
tr_A6L9E7 VAWDWWNDW.NLYG..VDFRAGINNPTYKYYIDFAAEHGIEYVILDEGWAVNLQADMMQV
tr_R6WPH8 VAWDWWNDW.NIYD..VDFRAGINTETYKYYIDFAAEHGIEYVILDEGWSVNMAADLMQV
tr_D4IM97 VAWDWWNTW.NLYG..VDFKSGINNETYKYYIDFAADHGIEYVILDEGWAVNKKADLFQV
tr_R6KSN5 VAWEWWNSC.GLSG..VDFKAGVNNVTYKAYIDFASKHGIEYVILDEGWAVNLKADLFQV
tr_D6CY08 VAWDWWNDW.NLYN..VDFRAGINNETYKYYIDFASKFGIEYVILDEGWAVPGKADLFEV
tr_A5FA11 VAWDWWNDW.NIYN..VDFKAGINTQTYKYYIDFASKNKVEYVVLDEGWSV..ETDIMKH
tr_E4T1A6 AAWDWWNNW.NITH..VDFKAGINTPTYKYYIDFASANKISYIILDEGWSV..EGDLSKV
tr_R7EPT5 VAWDWWNTC.NLTG..VDFKAGMNTPTYKKYIDFAAENQLEYIIIDEGWSGK.ESLMEDI
tr_A6KZF1 VAWDWWNDW.NLKG..VPFKAGINMDTYKYYIDFASRNGLEFIVLDEGWYDPKSGDMLTV
tr_A4ASM2 VAWDWYNAN.NIYG..VDFEAGVNDETYKYFIDFASNNGIEYVILDEGWTKS.TTEILED
tr_B1ZN91 VAWDWWNAN.NVFG..VDFKSGVNTATYKYYIDFAAANGLDYVVLDEGWYK..LGNVLDV
tr_C7PTT2 VQWDWWHYN.NIYD..VDFRAGINNETYKYYIDFAAKYGIEYVLLDEGWCD..TRDLMKL
tr_A3QFB6 IAWDWYNEN.NLKG..VDFKAGINTETYQYYADFAADFGIENILIDDGWSS..QEDVLNV
tr_Q5L8Y9 VSWEWWHDA.RLYG..VDFRSGFNMDSYKYYIDFASKFGIPYIIMDEGWAKN.TRDPFTP
tr_Q8A2L1 VSWEWWNGA.TPYGQDVTFKAGCNLETYKYFIDFASKFGIPYIIMDEGWAKS.TRDPYTP
tr_D5BGL4 VSWEWWNGA.SPYN..VDFVAGFNEETYKYFIDFASEFGIPYIIMDEGWAAS.TTDPYTP
tr_D2QSA9 CTDEWITNI.NLYN..VPFRAGVNTETYKYYVDFAKRFGFDRILLDAGWSD..NNDLFKI
tr_C6W680 ATDEWIITS.NIFN..VPFKSGINTATYKYYIDFAKRFGFERIMMDAGWSD..NNDLFKI
tr_R5UY73 VLWDWWNDR.NIYH..INFKSGINTDTYLYLVDYAAKHHIEYVLIDEGWSA..RNDLLTL
tr_E8V2F0 ASWNWWVDNVNGKG..ESGNKEFTTETMKYYVDFAAKSGFPYFFLDAGWA...SQDITKT
OWNID_5xfmA GTWSWILWQ.....DGSINYDDQVRYIDFASAMGYEYALIDNWWDT.....R
tr_A6KZ33 GTWSWILWQ.....DGSINFDDQKKYVDLAAAMGYEYVLIDNWWDT.....N
tr_R5VXS7 STWSWIMWQ.....DPSINYDDQIRFIDLASEMGYEYTLIDGGWEK.....N
tr_R5VWB3 GTWSWIIAG.....DASMNMDDQKRYVDFCAAMGYRTVLVDALWDT.....Q
tr_D6D0J3 GSWSWIIGM.....DGSTNYKEQLRYIDFSAAMGYQSVLVDALWDK.....Q
tr_A5FD29 GSWSWIIKM.....DNNTTFPVQKQYIDFSARMGYETILVDALWDT.....Q
tr_R5LYV6 STWSWILWQ.....DASCNYKDQQTFIDLAAAMGYEYILIDALWDK.....Q
tr_R6B488 YTWSWLIWQ.....DGSINYDDQVKMIDVAAAQGYEAILVDAWWDK.....Q
tr_R6C1F0 YAWSWLIWQ.....DESCNYNDQKLFIDLAATMGYEYVLVDALWDT.....Q
tr_A6KWX6 VSWIYWAYN.....HGSKDYQIVKKYIDFAADFDLPYMLIDWEWDA.....M
tr_R5V6Z4 VSWIYWAYN.....HGSRDYQIIKKYIDFAADFHLPYMLIDAEWDE.....M
tr_Q8A1J7 ASWVYWAHN.....HGSKNFKIVKDYIDLAVDMKWPYSLIDAEWNE.....M
tr_R6KK91 ASWSWWSNG.....GTPRDYKAQLKYVDLSAEMGWEYMLIDAGWQN.....M


α5
β25
α6
α7
PDBID_3a24A TT
350
360
370
380
PDBID_3a24A VKEIDLKELVDYAASKNVGIILWAGYHA.....FER..DMENVCRHYAEMGVK
tr_A6L9E7 IPEIDLQELVDYGKSKNVGIILWAGYWA.....FAR..DMENVVKHYSDMGVK
tr_R6WPH8 IPEIDIPELVNYGKSKNVGIILWAGYHA.....FDR..DMEKVVKHYSNMGVK
tr_D4IM97 VPEIDLPELVAYGKERGVGIVLWAGFYA.....TER..DLERVCKHYSEMGIK
tr_R6KSN5 VPEIDLKELVAYADSKHVGLILWAGYYA.....FER..DLERICKHYSELGIK
tr_D6CY08 IPEIDLKELISYAKSKNVDLILWAGYRA.....FEK..DMDRVCKHYAAMGIK
tr_A5FA11 NPNVDLEALIAYAKERNVGIILWASWMA.....LHE..NIDGVFDNYAKLGVK
tr_E4T1A6 SPAINLKEIVDYGKQKHVDVILWATWSN.....IAR..QMDEIFPQYSKMGVK
tr_R7EPT5 SPEINLEELVAYGNQKGIGIILWASWRNLTGGGD.....ISP..AVEQVISHYAQMGIK
tr_A6KZF1 IPELDLPELIAYGKSKGVEIVLWTVFNV.....LDS..QLEAACKKYADMGIK
tr_A4ASM2 NENIDVPGLIKYAKSKNVEIILWVLWKP.....LYE..DPEKILSLYASWGAV
tr_B1ZN91 VPEIDVPELVAYGRQKNVSIVLWVVWKT.....LED..QLQPALDQFAQWGVR
tr_C7PTT2 TPGIDVKELVDYAKAKNVNILLWSSWLV.....LDK..QLDMALDKFREWGVK
tr_A3QFB6 LPGIDMQAIIAHAKARGVGVQLWVPYSA.....LDK..NLEEAFKLYAKWGIN
tr_Q5L8Y9 NPTINLTELIKYGKDRNVKIVLWLPWLT.....VEN..HFD.LFKTFADWGIA
tr_Q8A2L1 NPDVDLHELIRYGKEKNVGIVLWLTWLT.....VEK..NFD.LFKTFNEWGIK
tr_D5BGL4 NPDIDLFELIEYGKKRNVKIVLWLTWLT.....VEH..HFD.LFEKFAEWGIA
tr_D2QSA9 TPGMNIPELVAYAHSQGIKLSMWTLAAT.....LDR..QLEPAMKQFQSWGVD
tr_C6W680 NPDIDMDEIVAYAKSKGVGISMWTLALT.....LDK..QLEPALKQFSKWGVD
tr_R5UY73 NPDVDIPRICEYATEKGVGIQLWSKWVN.....IMR..QMDEAFAQFSKWGVK
tr_E8V2F0 NGRVDVPELVRYAAARNVKVWIWLYSEA.....VMK..QMKEAFPVYEQWGVA
OWNID_5xfmA IGHQRMKSLVEYARDKGVELFLWYSSSGYWNDIEQGPVNRMDNAIIRKREMKWLQSLGVK
tr_A6KZ33 IGRERMKDFIDYAHSKNVDVFLWYSSSGYWNDIVQGPTNYMDNPIIRKKEMKWLHNIGVK
tr_R5VXS7 IGRDRMEELFKYAHQKNVDVFVWYNSNGYWNDAPQDAKQCMSNSVARKKEMKWLQKCGVK
tr_R5VWB3 VGYDKIEELAAYAKQKNVGLYLWYNSNGYWNDAPQGPRGKMSNIIERRKEMKWMQRIGIK
tr_D6D0J3 IGRDKIEELAKYGKDKGVALYLWYNSNGYWNDAPQTPRGIMDNAIARRKEMKWMQSIGIR
tr_A5FD29 IGKDKIAELAQYGTEKGVGLYLWYNSNGYWNDAPQGPRGMMDNSIIRRKEMAWLKEIGIK
tr_R5LYV6 IGYENMPSLIAYAQSKGVDVLLWYNSNGNWNDAPQGPKHRLDSAPARQKEMAWMKTLGVK
tr_R6B488 IGRKRIEELSKYAKSKKVSLMLWYNSNGFENDAPQTPRQIMNNSIARKKEMAWMKKIGVV
tr_R6C1F0 IGRDKIAELSKYAQSKNVSLMLWYNSNGVANDAPQGPRGIMDNIVARKKEMAWMKSIGIK
tr_A6KWX6 GNGGNLNDAMNYCKEKNVTPLIWYNSSTAW..CDPTPLYRLNTPEVRDKEFEWLKEIGIQ
tr_R5V6Z4 GNGGNVEDALAYAKQKNIAPLLWYNSSTAW..CAFGPLYRLNKPEMRQKEFAWLESMGVK
tr_Q8A1J7 SNGGDIEDLITYSLSRNIKPMIWYNSSTAW..MGPGPLYRLNSKENRIKEYTWLQKSGVV
tr_R6KK91 GNGGTMEDVVKYAQQKGVGVWLWYHSGAGREKDSVSTHRLMSDPELRRAEMKRISEVGVK
β26
α8
β27
α9
β28
PDBID_3a24A
390
400
410
420
430
440
PDBID_3a24A GFKVDFMDRDDQEMTAFNYRAAEMCAKYKLILDLHGTHKPAGLNRTYPNVLNFEGVNGLE
tr_A6L9E7 GFKVDFLDRDDQEMVNFVYKASEVCAKYKMLVDFHGVFKPTGLQRTYPNVLNYEGVNGLE
tr_R6WPH8 GFKVDFMDRDDQEIVDFLYRGAETCAKYKMMVDYHGICKPTGLQRTYPNVINYEGVHGLE
tr_D4IM97 GFKVDFLDRDDQKIADFTYRLAETAAKYRLFLDMHGYHKPAGIQRTYPNVLNFEGVFGLE
tr_R6KSN5 GFKVDFMDRDDQAMVDFHYRGAEIAAKYHLMLDYHGTYKPTGMNRTYPNVINFEGVHGLE
tr_D6CY08 GFKIDFMDRDDQQVVEFNRKAAETGAKYKLLIDLHGTFKPTGLQRTYPNTINFEGVHGLE
tr_A5FA11 GFKVDFIDRDDAKMVNSVYEIAQKAANHKLIIDFHGMYKPTGIQRTYPNILNFEGVKGLE
tr_E4T1A6 GFKIDFFDRDDQVVVASTYAIAKKAAEYHLMVDYHGIYKPTGIQKTYPNVIGFEGVKGME
tr_R7EPT5 GFKIDFFDRDDQLAISSTEQLAACAAKHHMLLDLHGL.KPFGIQRAYPNILNFEGVKGLE
tr_A6KZF1 GFKVDFLDRDDQTAVEMVYRIAEMTARYKLTLDLHGIYKPTGINRTYPHIINFESVFGME
tr_A4ASM2 GIKTDFMQRSDQYVVESYEKIAEIAAKYKLLVDYHGAFKPAGVERMWPNIINYEGVRGNE
tr_B1ZN91 GIKVDFMQRDDQKVMSHYHRVSAEAAKRRLLVDFHGAIRPALLTRTYPNLITTEGVRGNE
tr_C7PTT2 GIKVDFMQRDDQDMVNYYEKVSKAAAARQLMVDFHGAYKPTGWLRTYPNVLTSEGVLGNE
tr_A3QFB6 GVKIDFMDSDSIKRVDFYWRAAAMAAKYKLMVNFHGAYKPAGIHRTYPNVMTREGVRGLE
tr_Q5L8Y9 GVKIDFMDRSDQWMVNYYERVAKEAAKHKLFVDFHGAFKPAGLERKYPNVLSYEGVLGME
tr_Q8A2L1 GVKIDFMDRSDQWMVNYYERVAQEAARHHLFVDFHGSFKPAGLEYKYPNVLSYEGVRGME
tr_D5BGL4 GIKIDFMDRSDQWMVHFYERVAKEAAKHELFVDFHGSFKPAGLEYKYPNILSYEGVLGME
tr_D2QSA9 FVMTDFMDRDDQKMVNFYTRVAEACAKAKIMIMYHGAYKPAGFERTYPNAITRESVLGSE
tr_C6W680 FVMTDFINRDDQKAVQFFYKIAEACARHKIMLMFHGAFKPAGFNRTFPHAVARESVLGSE
tr_R5UY73 GVKIDFMDRNDAKMVNFYEQVAIKATQYKLLVDFHGSYPNEGMRRKYPNLMTREGVIGLE
tr_E8V2F0 GMKIDFVNRDDQEGIQFFYDVARYAAEHHLMVDFHGCHTPWGLMRAYPNVMGYESVLGLE
OWNID_5xfmA GIKVDFGGDKQETMRLYEDILSDADDHGLMVIFHGCTLPRGWERMYPNYVGSEAVLASE
tr_A6KZ33 GIKVDFGGDKQETMRLYEAILSDADDNGLMVIFHGCTLPRGWERMYPNYVGSEAVLASE
tr_R5VXS7 GLKVDFGGSKQQMMGLYEDILTDANEYGLMIIFHGCTIPRGWERMYPNYVGSEAVLASE
tr_R5VWB3 GIKVDFIGSDKQVTMQLYEDILADANDYGLLVIFHGCTIPRGWERMYPNYAASEAVLASE
tr_D6D0J3 GIKVDFGGDKQMTMQLYEDILSDANEYGLLVIFHGCTLPRGWERMYPNFASSEAVLASE
tr_A5FD29 GIKVDFGGDKQVTMKLYEDILNDANDFGLMVIFHGCTLPRGWERMYPNYASSEAVLASE
tr_R5LYV6 GIKVDFGGDKQTTMKLYEDILTDANQYGIAVVFHGCTLPRGWERMYPNHMSSEAALVSE
tr_R6B488 GIKVDFGGDKQETMKLYEDILSDANDYGLEVIFHGCTMPRGWERMYPNYVSSEAALASE
tr_R6C1F0 GIKVDFGGDKQHTMQQYEDILSDANDYGIQVIFHGCTLPRGWERMYPNYVSSEAVLASE
tr_A6KWX6 GVKIDFGGDSIGTMNYYIDLLEDAARHKLLVNFHGAAIPRGWQRTYPNLMSVEAVYGSE
tr_R5V6Z4 GIKVDFFSQDSISTMNYFVDILEDAAKYHLMVNFHGAPIPRGWQRTYPHMMSVEGVYGAE
tr_Q8A1J7 GVKIDFFPDDYSSNMDYYIDLLEDAAKYKLMVNFHGATIPRGWQRTYPHLMTMEAVYGAE
tr_R6KK91 GIKVDFFDTDKQRIIQLYPALLKDAADNHLLVDLHGATLPRGFERTYPNMLTTEAIRGAE


η3
α10
η4
α11
β29
PDBID_3a24A
450
460
470
480
490
PDBID_3a24A QMKW....SS.PSVDQVKYDVMIPFIRQVSGPMDYTQGAMRNASK.......GNYY..PC
tr_A6L9E7 QMKW....SSEKEFDMVTYDVTIPFIRMIAGPMDYTQGAMRNAAR.......GNHR..SV
tr_R6WPH8 NMKW....VA.STYDMPLYDVTIPFVRMVAGPMDYTQGAMRNAIR.......KNYA..PI
tr_D4IM97 QMKW....TS.EKTDMVTYDVTIPFIRMLAGPMDYTQGAMRNATR.......RNFR..AV
tr_R6KSN5 QLKF....SGSENVDQVTYDVTMPFIRMIAGPVDYTQGAMKNGNK.......RNFR..AV
tr_D6CY08 EMKW....AE.PGTDQVIYDVTAPFIRMVAGPLDYTQGAMNNVIM.......KNFH..AV
tr_A5FA11 NNKW......TPNDDVPLYDTTIPFIRMMAGPMDYTPGAMRNATK.......SEFK..PS
tr_E4T1A6 NYKW.......ATEDQPRYAVSIPFIRMMAGPMDYTPGAMRNAIK.......ANYH..PV
tr_R7EPT5 NSKWEPIVNGAPLHDFPRYDVTAPYLRQLAGPMDYTPGAMVNTTR.......SLFR..SV
tr_A6KZF1 EVKW....TD.IKNNMPLYDVTFPYIRMMAGPVDYTPGAMRNATK.......ADWR..AM
tr_A4ASM2 NNKW.......SSAITPEHNVTLPFVRMAAGPFDYTPGAMINMNKIDYETGGSNYA..PL
tr_B1ZN91 WNKW.......STHITPEHTATLPFTRMMMGPMDFTPGATLNANR.......TGFA..AN
tr_C7PTT2 ISKF.......EDWITPSHTTTIPFIRMAAGPMDFTPGGMQNVQR.......DSWS..AI
tr_A3QFB6 NNKW........SEVTSTHNLHLPFIRMIAGPLDYTPGAMANAQQ.......KNFN..KV
tr_Q5L8Y9 QGGN..........CKPENSIYLPFMRNAVGPMDFTPGSMISAQP.......EDNR..ST
tr_Q8A2L1 QMGG..........CKPENSIYLPFMRNAVGPMDYTPGAMISMQP.......NIYR..SE
tr_D5BGL4 QMGR..........ATPDNSLFLPFMRNAVGAMDYTPGAMISMQP.......EFYY..SR
tr_D2QSA9 YNIW.......STKATPDHDVLLPFIRMVSGPMDYEPGLLSNATK.......EQFR..PI
tr_C6W680 YNMW.......SAKATPEHNVLLAFIRMLGGPLDYEPGMLNNATP.......KGFH..PM
tr_R5UY73 YNKW.......SKRATVTHDVIIPYLRMWVGPMDYTPGAMLNAHP.......ETFY..EN
tr_E8V2F0 NNKV.......GRRDSPVDRAVFAWTRLLVGPMDYTAGGFDNQTE.......DSFV..AR
OWNID_5xfmA NMVF....NQHFCDEEAFNTCLHPFIRNTVGSMEFGGCLLNKRLN.......RNND..GG
tr_A6KZ33 NLIF....QQHFCDEEAFNACLHPFIRNTIGCMEFGGTFLNKRLN.......RNND..GG
tr_R5VXS7 NLIF....NQHFDDMEAYNACLHPFIRNTIGCMEFGGTLLNKRHN.......RTND..GG
tr_R5VWB3 NMNF....SQGSCDAEAFNACIHPFVRNTIGSMDFGGSTLNKYYN.......AKNEL.GG
tr_D6D0J3 NLHF....SQGSCDNEAFNATLHPFIRNTVGSMDFGGSALNKYYN.......ADNAP.RG
tr_A5FD29 NLHF....GQASCDKEAMNASIHPFIRNTVGSMDFGGSALNKFYN.......SDNTPNKG
tr_R5LYV6 NLVF....GQYHADKEAQRSTLYPFIRNAVSAMDFGGVFFNKHFS.......KDGV..SG
tr_R6B488 NVYF....TDYHAKKEAFEMTMHPFSRNAVGSFDWGGVMMNKYFS.......KDNK..SR
tr_R6C1F0 NVFF....SDYHAKQEAFELTMHPFCRNAVAAMDWGGTIMNKYLS.......KDNK..SR
tr_A6KWX6 WYNN....KPTLTDNAAWHNCTLPFTRNVIGPMDYTPCTFTNSQH..............
tr_R5V6Z4 WYNN....KETLTDSAAWHNCTLPFTRNVVGPMDYTPCTFTDSQH..............
tr_Q8A1J7 WYNN....KPILTDRAAEHNVTLPFTRNVIGSMDYTPGTFSDSQH..............
tr_R6KK91 TLGR.....QERCDQAAKHNATVPFTRNVVGSMDYTPVTFSDKVR.......QGVP....
catalytic domain
C-terminal domain
β30
α12
β31
α13
α14
β32
PDBID_3a24A
500
510
520
530
540
550
PDBID_3a24A YSEPMSQGTRCRQLALYVVFESPFNMLCDTPSNYMRE.PE.STAFIAEIPTVWDESIVLD
tr_A6L9E7 NSEPMSQGTRCRQLATYVIFESPFNMLCDAPSNYRRE.KE.CTEFISNIPTIWDETVSLA
tr_R6WPH8 YTEPMSQGTRCHQLATYVIFESPLNMLCDNPSNYNRE.PE.CTEFIANIPTVWEKTVALD
tr_D4IM97 GSEAMSQGTRCRQLAEYVIFESPFNMLCDSPSNYMRE.PE.CTEFIASVPTTWDETVGLD
tr_R6KSN5 NEEPMSMGTRCRQLAEYVIFEAPLSMLCDSPVLYERE.SE.CTSYISDIPTVWDETRALN
tr_D6CY08 YTEPMSQGTRCRQLALYTIFDSPINMLCDAPTNYLKE.EE.CTKFIAAIPTVWNQTLPIS
tr_A5FA11 HSNPMSQGTRCHQLALYTIFEAPLQMMADSPTAFMKE.QE.STDFIAKVPTTFDETAALD
tr_E4T1A6 SSAPMSMGTRCNQLAMYVVFEAPFQMLADNPTIYMKE.QE.CTDFITGVPTTTDETVALD
tr_R7EPT5 NDHPMSQGTRVHQMAMYTVFEAPLQMLADSPSKYMKE.QE.CTDFIAKVPTTFDETIALD
tr_A6KZF1 YYTPASMGTRCHQLAAYIVHDSPFTMLCDAPTNYLNE.QE.CVDFIASLPVEVDSTFIAS
tr_A4ASM2 FTRPMAFGTRAHQVAMYVVFEAPVQMLCESPTIYYKE.QE.TVDFITQIPTVWDETVVLE
tr_B1ZN91 FARPSSQGTRCHQLALYVVFESPLQMLCDSPSHYLRE.PD.MMEFLRSVPTEWDETRVLD
tr_C7PTT2 PAEPLTLGTRSNQLAMYVVYESPLQMLCDLPTHYYHE.PV.AMDFLQTVPTVWQQTIPLQ
tr_A3QFB6 WSRPMSLTTRAQQIAMYVIYESPLQMLADTPTSYRAE.PE.IPRFIAAIPTVWDDFKVLH
tr_Q5L8Y9 RANAMGSGTRAFQMALFIIFESGLQMLADNPVYYYRE.LP.CTEFITSVPVTWDETKVLY
tr_Q8A2L1 RPNSASIGTRAYQMALFVIFESGLQMLADNPTLYYRN.ED.CTRFITQVPVTWDETVVLE
tr_D5BGL4 RPNSAAIGTRAQQMAFFVIFESGLQMLADNPTNYFKN.EE.CTQFITDVPVTWDQTIALE
tr_D2QSA9 SGQVMSQGTRCHQLAMFLVYDSPISIFSGNPSQALLE.PR.FMEFLGSVPSAFDETKVLD
tr_C6W680 PENVMSLTTRTQQLAMFVVYESPIQFFAGNPSQGLKE.PG.FMELLGSIPTVWDTTRILD
tr_R5UY73 QHEPMSQGTRSHQLAMYVVYESPLQMISDSPTKYDKN.LR.SFEFIKSIPTTWDETVPLE
tr_E8V2F0 NATPQVMGTRAQQLALYVVYENPIPMVSDSPQNYDGE.AAPAFQFLKDVPTTWDETHVIG
OWNID_5xfmA ...TTRRTTDVFQLATTVLLQNPVQNFALAPNNLKDVPAV.CMDFMKRVPTTWDETRFVD
tr_A6KZ33 ...NIRKTTDVFQLATAVLFQNPIQNFALAPNNLTDAPQV.CLDFMKQVPTTWDETRFID
tr_R5VXS7 ...TIRKTTDVFQLATAVLYQNPIQNFALAPNNLTDAPAL.AIDFMKQVPTTWDETVFID
tr_R5VWB3 ...RMRVTSDVFALATAVLFQSPVQHFALAPNNLTDAPAW.AVDFMKKVPTTWDEVKFID
tr_D6D0J3 ...SRRVTSDVYALATAVLFQSPVQHFALAPNNLTDAPSW.AIDFMKEVPTTWDEVRFID
tr_A5FD29 ...SKRITSDVFALATAVLFQSGVQHFALAPNNLTDAPDW.AVKFMKEVPTTWDEVRFLD
tr_R5LYV6 ...TLRKTTDAFQIATSVLYQSGIQHFGITPNNLTEQPEF.ILDFMKKVPTVWDETRFID
tr_R6B488 ...HQRNTSDIFEMATAVTNQSSVNCVCLYPNNLQDVPQW.ELDWLKAVPTAWDDVKFIA
tr_R6C1F0 ...HRRYTTDVFEMASAIMNQAAIQCIAIYPNNLSELPQH.EIDFLKSVPTTWDETKFIA
tr_A6KWX6 ....PHITTDAHELALLVIFESALQHLADRPEGYRQQPVE.VQNFIRYLPVAWEDTRLLS
tr_R5V6Z4 ....PHITSNGHELALLVAFESALQHLADRPEGYHAQPAE.IRSFISSLPTVWDDTKLLS
tr_Q8A1J7 ....PHITTCGHELALSVVFESALQHMPDRPEVYSSLPRK.VREFLSGLPTAWDDTKLLC
tr_R6KK91 ...AIRKTTMGHQLALAVVFESGFQCFADKAESYLSLPEQ.PKQFLKEVPAAWDESVLLA

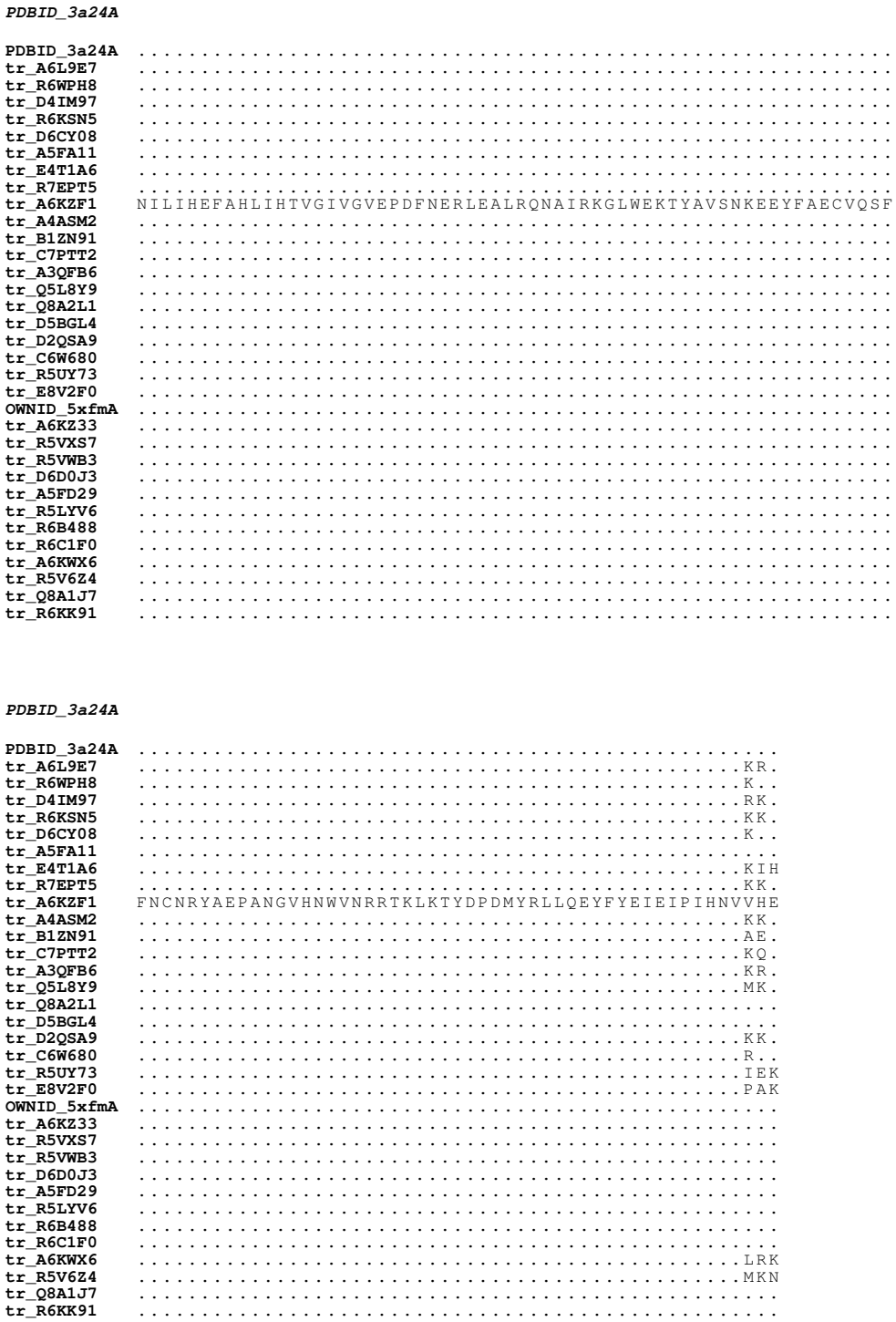

β33
β34
β35
PDBID_3a24A
560
570
580
590
C-terminal domain
β36
β37
β38
β39
600
610
620
630
640

PDBID_3a24A

```
PDBID_3a24A  ............................................................
tr_A6L9E7    ............................................................
tr_R6WPH8    ............................................................
tr_D4IM97    ............................................................
tr_R6KSN5    ............................................................
tr_D6CY08    ............................................................
tr_A5FA11    ............................................................
tr_E4T1A6    ............................................................
tr_R7EPT5    ............................................................
tr_A6KZF1    KNIPSFYQKYIETEGLYVTSSGKVSDEALLKACDIISLMLAKRPDVKAHMVKKGCHVMII
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    GKDEETCDLPEFAHICNCEDSIKYWNWRARGFGGAPEDEFSSSCGEENLLALPQDKYVGE
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    ............................................................
```



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://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi).