Running title: Primary Structure of *Bacillus* sp. AAH-31 α-Amylase

A Thermophilic Alkalophilic α-Amylase from *Bacillus* sp. AAH-31 Shows a Novel Domain Organization among Glycoside Hydrolase Family 13 Enzymes

Wataru SABURI,[1,*] Naoki MORIMOTO,[2,*] Atsushi MUKAI,[1] Dae Hoon KIM,[1] Toshihiko TAKEHANA,[2] Seiji KOIKE,[2] Hirokazu MATSUI,[1] and Haruhide MORI[1,†]

[1]*Research Faculty of Agriculture, Hokkaido University, N-9, W-9, Sapporo 060-8589, Japan*

[2]*Adeka Corporation, 7-2-34 Higashi-ogu, Arakawa-ku, Tokyo 116-8553, Japan*

Received April 3, 2013; Accepted May 30, 2013

[†]To whom correspondence should be addressed. Tel/Fax: +81-11-706-2497; E-mail: hmori@chem.agr.hokudai.ac.jp

*These authors contributed equally to this work.

Abbreviations: GH, glycoside hydrolase; AmyL, *Bacillus* sp. AAH-31 α-amylase; PVDF, polyvinylidene difluoride; CD, cyclodextrin; TVAI, *Thermoactinomyces vulgaris* R-47 α-amylase I; TVAII, *T. vulgaris* R-47 α-amylase II; CBM20, carbohydrate-binding module family 20; rAmyL, recombinant AmyL; pNPG5, *p*-nitrophenyl α-maltopentaoside; GA SBD, starch-binding domain of *Aspergillus niger* glucoamylase

α-Amylases (EC 3.2.1.1) hydrolyze internal α-1,4-glucosidic linkages of starch and related glucans. *Bacillus* sp. AAH-31 produces an alkalophilic thermophilic α-amylase (AmyL) of higher molecular mass, 91 kDa, than typical bacterial α-amylases. In this study, the AmyL gene was cloned to determine its primary structure, and the recombinant enzyme, produced in *Escherichia coli*, was characterized. AmyL shows no hydrolytic activity towards pullulan, but the central region of AmyL (Gly395-Asp684) was similar to neopullulanase-like α-amylases. In contrast to known neopullulanase-like α-amylases, the N-terminal region (Gln29-Phe102) of AmyL was similar to carbohydrate-binding module family 20 (CBM20), which is involved in the binding of enzymes to starch granules. Recombinant AmyL showed more than 95% of its maximum activity in a pH range of 8.2-10.5, and was stable below 65°C and from pH 6.4 to 11.9. The k_{cat} values for soluble starch, γ-cyclodextrin, and maltotriose were 103 s^{-1}, 67.6 s^{-1}, and 5.33 s^{-1}, respectively, and the K_m values were 0.100 mg/mL, 0.348 mM, and 2.06 mM, respectively. Recombinant AmyL did not bind to starch granules. But the substitution of Trp45 and Trp84, conserved in site 1 of CBM20, with Ala reduced affinity to soluble starch, while the mutations did not affect affinity for oligosaccharides. Substitution of Trp61, conserved in site 2 of CBM20, with Ala enhanced hydrolytic activity towards soluble starch, indicating that site 2 of AmyL does not contribute to binding to soluble long-chain substrates.

Key words: α-amylase; neopullulanase; glycoside hydrolase family 13; carbohydrate-binding module family 20; starch-binding domain

α-Amylases (EC 3.2.1.1), widely distributed in living organisms including animals, plants, and microorganisms, catalyze the hydrolysis of internal α-1,4-glucosidic linkages of starch, glycogen, and related glucans with net retention of the anomeric configuration. Fungal and bacterial α-amylases are utilized in various industries, including the production of isomerized sugar (liquefaction and saccharification of starch granules), bakery applications, and the de-sizing of textiles and paper. Alkalophilic liquefying enzymes are particularly useful as additive compounds for dishwashing and laundry detergents to remove food residues on dishes and food stains from clothes.[1)]

According to a sequence-based classification of glycoside hydrolases (GHs),[2)] α-amylases are classified into GH families 13, 57, and 119, while most enzymes are found in GH family 13. GH family 13 is the largest of the GH families. It is composed of various retaining GHs and glycosyltransferases that act on α-glucans such as starch, maltooligosaccharides, and sucrose.[3)] Enzymes of this family show three-dimensional structures similar to enzymes belonging to GH families 70 and 77, and hence these families are categorized into clan GH-H, regarded as the α-amylase family.

GH family 13 enzymes share three domains (A, B, and C).[3)] Domain A is the catalytic domain formed by a $(\beta/\alpha)_8$-barrel. Domain B is a long loop connecting the third β-strand and the third α-helix, and a substrate binding cleft is formed at the interface of domains A and B. Domain C, following domain A, is formed by an antiparallel β-sheet. GH family 13 enzymes have been further classified into 40 subfamilies based on amino acid sequences.[4)] Most subfamilies contain enzymes that utilize the same substrate and produce the same product. α-Amylases are found in subfamilies 1, 2, 5, 6, 7, 15, 19, 20, 24, 27, 28, and 32.

Bacillus sp. AAH-31, first isolated from Japanese paddy soil, produces extracellular liquefying α-amylase (AmyL), which is highly stable at high pH and temperatures.[5)] Many alkaline α-amylases have been identified in various alkalophilic bacteria,[6-11)] but AmyL is more stable at high temperature than the known alkalophilic enzymes. Furthermore, this enzyme is stable in the presence of chelating reagents, including

EDTA, nitrilotriacetic acid, and sodium tripolyphosphate. These properties of AmyL are useful in automatic dishwashing detergents, because an automatic dishwashing machine generally operates at high pH at temperatures above 60°C. On SDS-PAGE analysis, the molecular mass of AmyL was estimated to be 91 kDa, considerably higher than those of the typical bacterial α-amylases of GH family 13 (approximately 60 kDa) and close to enzymes possessing starch-binding domains.[11,12] In the present study, the gene encoding AmyL was cloned to elucidate its primary structure, and the recombinant enzyme, produced in *Escherichia coli*, was characterized. Furthermore, a site-directed mutational study at the starch-binding site, which was identified through sequence analysis, was carried out.

Materials and Methods

Analysis of N-terminal and partial internal amino acid sequences. Ten μg of AmyL purified from the culture supernatant of *Bacillus* sp. AAH-31[5] was further separated by SDS-PAGE and transferred to a polyvinylidene difluoride (PVDF) membrane (Sequi-Blot PVDF Membrane, Bio-Rad, Hercules, CA) by electroblotting with a semi-dry blotting apparatus.[13] The band of AmyL, detected by staining with Coomassie Brilliant Blue, was cut off from the membrane and subjected to N-terminal sequence analysis with a Procise 492HT protein sequencer (Perkin Elmer, Waltham, MA). To analyze partial internal amino acid sequences of AmyL, the enzyme was digested by endoproteinase Asp-N (Takara Bio, Otsu, Japan). Twenty-five μg of AmyL was transferred to a PVDF membrane as described above, and the band of AmyL was incubated in 100 μL of 50 mM sodium phosphate buffer (pH 8.0) containing 0.65 ng/μL of endoproteinase Asp-N for 4 h at 37°C. Ten μL of the resulting supernatant was subjected to separation by HPLC under the following conditions: column, Capcell Pak C8 (ϕ4.6 × 250 mm, Shiseido, Tokyo); column temperature, room temperature; elution, 0-60% acetonitrile; flow rate, 0.5 mL/min; detection, absorbance at 280 nm. The N-terminal amino acid sequences of the purified peptides were analyzed as described

above.

Gene cloning of AmyL *and construction of its expression plasmid.* To obtain partial *AmyL*, PCR was carried out with LA Taq DNA polymerase (Takara Bio), the genomic DNA of *Bacillus* sp. AAH-31, and primers designed from the N-terminal amino acid sequence and an internal amino acid sequence (P2, described below) of AmyL, S: 5′-CARGAYTAYGARAAYATHGT-3′ (sense), and R: 5′ -TCYTCIACRTAYTGYTCRTC-3′ (antisense). The amplified DNA fragment was cloned into the *Eco*RV site of the pBluescript II SK (+) vector (Stratagene, La Jolla, CA), and sequenced with an ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA). Two μg of the genomic DNA digested with *Hin*dIII (Takara Bio) was cyclized through self-ligation with T4 DNA Ligase (Takara Bio), and used as the template for inverse PCR. The regions upstream and downstream of the section obtained by the PCR, as described above, were amplified with cyclized genomic DNA as template. Primers, HS1: 5′-ACCATCGGCTCCAGTTTG-3′ (sense) and HR1: 5′-TGGATGATTGTTTGAACTCCAG-3′ (antisense), were used for the upstream region, and primers, HS2: 5′-ACCAGTGGCAGATTGAGC-3′ (sense) and HR2: 5′-GGCACTTTCTACTTTGGAAAG-3′ (antisense), for the downstream region. The further downstream region was amplified by inverse PCR, in which the genomic DNA cyclized after digestion by *Bam*HI was used as template. The primers used for this PCR were BS: 5′-TGTCCCAAAGTGAGGACCAT-3′ (sense) and BR: 5′-CTCTGTTCCGTAATAGATGGC-3′ (antisense).

AmyL was inserted into pET-21a vector (Novagen, Darmstadt, Germany) to overexpress it in *E. coli*. The DNA fragment amplified by PCR was cloned into pET-21a *via* the *Nde*I and *Xho*I sites, using Primestar HS DNA polymerase (Takara Bio), the genomic DNA of *Bacillus* sp. AAH-31, and primers 5′-ACGCATATGCAGGATTATGAGAACATTGTCCTAAGA-3′ (sense, *Nde*I site underlined) and 5′-GTTCCCCTCGAGTTTGTTAATGTTGCCCAGGACC-3′

(antisense, *Xho*I site underlined).

Preparation of recombinant AmyL. The transformant of *E. coli* BL21 (DE3) harboring the expression plasmid for *AmyL* was cultured at 37°C in 300 mL of Luria-Bertani medium supplemented with 100 μg/mL of ampicillin until A_{600} reached 0.5, and 0.3 mL of 0.1 M isopropyl 1-thio-β-D-galactopyranoside (Wako Pure Chemical Industries, Osaka, Japan) was added to induce production of the recombinant protein. The induction culture was carried out at the same temperature for 3 h. Bacterial cells harvested by centrifugation were suspended in 10 mM Tris-HCl buffer (pH 7.0) containing 0.5 M NaCl and disrupted by sonication. After the cell debris was removed by centrifugation, the cell-free extract was applied to a Ni-chelating column by Chelating Sepharose Fast Flow (GE Healthcare, Uppsala, Sweden). The column was thoroughly washed with 10 mM Tris-HCl buffer (pH 7.0) containing 0.5 M NaCl, and the adsorbed protein was eluted with 0.15 M imidazole in the same buffer. Highly purified fractions, confirmed by SDS-PAGE, were collected and dialyzed against 10 mM Tris-HCl buffer (pH 7.0). The protein concentration was determined from the concentrations of amino acids after complete acid hydrolysis of the purified enzyme with 6 *N* HCl at 110°C for 24 h. Amino acids were measured by the ninhydrin colorimetric method using a JLC-500/V (JEOL, Tokyo).[14)]

Standard enzyme assay. A reaction mixture (100 μL) consisting of an appropriate concentration of the enzyme, 100 mM Tris-HCl buffer (pH 9.5), 5 mg/mL of soluble starch (Nacalai Tesque, Kyoto, Japan), and 0.025 mg/mL of Triton X-100 was incubated at 50ºC for 10 min, and the resulting reducing sugar was measured by the dinitrosalicylic acid method,[15)] in which 0-5 mM maltose (Nacalai Tesque) was used as the standard reducing sugar. One U of enzyme activity was defined as the amount of enzyme producing 1 μmol of reducing sugar per min.

Effects of pH and temperature on the activity and stability of recombinant AmyL. Enzyme activities at various pH values were measured to determine the optimum pH of AmyL. One hundred mM Britton-Robinson buffer (pH 6.0-11.0, composed of a mixture of 100 mM acetate, phosphate, and glycine, pH adjusted with 5 *N* NaOH) was used as the reaction buffer. The other reaction conditions were same as for the standard enzyme assay.

To evaluate pH and temperature stabilities, the enzyme solution was incubated at various pHs and temperatures, and residual activity was measured by the standard enzyme assay. For the pH studies, the enzyme solution was incubated in 100 mM Britton-Robinson buffer (pH 4-13) at 4°C for 24 h. For heat treatment, the enzyme solution was incubated at 30-80°C for 15 min, and then cooled down immediately on ice. The enzyme was considered to be stable in those ranges of pH and temperature over which the enzyme maintained more than 90% of its original activity.

Kinetic parameters. The reaction velocities for the hydrolysis of various concentrations of soluble starch, γ-cyclodextrin (γ-CD), and maltotriose were measured. In the hydrolysis of soluble starch, a reaction mixture (500 μL) consisting of 1.07-2.14 nM enzyme, 100 mM Tris-HCl buffer (pH 9.5), 0.05-0.3 mg/mL of soluble starch, and 0.025 mg/mL of Triton X-100 was incubated at 50ºC. In the reaction to γ-CD, 0.1-1.0 mM γ-CD (Nacalai Tesque) was added to the reaction mixture. Fifty μL of the reaction mixture was removed every 2 min up to 10 min, and the reducing sugar produced was measured by the copper bicinchoninate acid method,[16] in which 0-300 μM maltose was used as the standard. Initial reaction velocities, determined from the slopes of the progress curves, at various substrate concentrations were fitted to the Michaelis-Menten equation by nonlinear regression with Grafit version 7.0.2 (Erithacus Software, West Sussex, UK). Data shown are mean ± standard deviation for three independent experiments.

For the analysis of maltotriose as substrate, a reaction mixture (100 μL) consisting of

49.5-94.5 nM enzyme, 100 mM Tris-HCl buffer (pH 9.5), 1-8 mM maltotriose (Wako Pure Chemical Industries), and 0.025 mg/mL of Triton X-100 was incubated at 50°C for 10 min, and heated at 100ºC for 10 min to terminate the reaction. After the addition of 200 μL of 2 M Tris-HCl buffer (pH 7.0), the D-glucose released was measured by the glucose-oxidase peroxidase method.[17] Kinetic parameters were determined as described above.

Estimation of the molecular mass of recombinant AmyL. To determine the molecular mass of recombinant AmyL, gel filtration column chromatography was carried out. A Superose 12 column (ϕ10 × 300 mm, GE Healthcare) was used, and column chromatography was performed in 10 mM Tris-HCl buffer (pH 8.5) containing 100 mM KCl at 0.5 mL/min. The eluted protein was detected by the absorbance at 280 nm. A set of gel filtration standards (Bio-Rad) was used to calibrate the molecular mass.

Analysis of the ction pattern of recombinant AmyL on p-*nitrophenyl α-maltopentaoside* (*pNPG5).* One mL of a reaction mixture consisting of 5.6 nM recombinant AmyL, 1 mM pNPG5 (Calbiochem, Bad Soden, Germany), and 20 mM Tris-HCl buffer (pH 9.5) was incubated at 50°C. At 2, 5, 10, and 15 min, 100 μL of the reaction mixture was removed and mixed with 10 μL of 5 M acetic acid to terminate the reaction. The reaction products were analyzed by HPLC under the following conditions: injection volume, 10 μL; column, YMC C18 Pack Pro (ϕ6.0 mm × 250 mm, YMC, Kyoto, Japan); column temperature, 40°C; eluent, 10% acetonitrile; flow rate, 1 mL/min; detection, absorbance at 313 nm.

Preparation of W45A, W61A, and W84A AmyL mutants. Site-directed mutagenesis was carried out with a Primestar Mutagenesis Basal Kit (Takara Bio). The expression plasmid for wild-type AmyL was used as template. The sequences of the primers used in the PCRs were as follows: for W45A,

5′-CTTGAC**GC**GAGTTCAAACAATCATCCA-3′ (sense, nucleotides mutated in boldface) and 5′-TGAACTC**GC**GTCAAGGGGGGCCAAGCT-3′ (antisense); for W61A, 5′-GGAACT**GC**GAAAAGCAATCCCATCCCT-3′ (sense) and 5′-GCTTTTC**GC**AGTTCCATCTGACTCATC-3′ (antisense); for W84A, 5′-GGGCAA**GC**GCTGCCGGGAGAGAATCTC-3′ (sense) and 5’-CGGCAGC**GC**TTGCCCATCCATCACATA-3′ (antisense). The expression plasmid of each of the mutant enzymes was introduced into *E. coli* BL21 (DE3), and the enzymes were prepared in the same fashion as the wild type.

Results

Cloning and sequence analysis of AmyL

The N-terminal and partial internal amino acid sequences of AmyL purified from the culture supernatant of *Bacillus* sp. AAH-31 were analyzed. The N-terminal sequence of AmyL was QDYENIVLRGSLAPLDW, and the sequences of the peptides derived from AmyL by Asn-N endoprotease digestion were P1, DWAK; P2, DEQYVE; and P3, DWSSNNHPLTK. From PCR using primers, based on the sequence of the N-terminal and P2 and genomic DNA as template, a 488 bp DNA fragment was obtained (Fig. 1). The sequence of P3 was confirmed in the deduced amino acid sequence of the DNA fragment. To obtain the regions upstream and downstream of the DNA fragment obtained, inverse PCRs were carried out successively, and a 3,533-bp portion of the DNA sequence, including the full-length of *AmyL* (2,466 bp) encoding 821 amino acid residues, was determined. The nucleotide sequence has been deposited in the DDBJ database under accession no. AB775769 (Fig. 1). No ATG codon was found in the upstream region encoding the N-terminal amino acid sequence of AmyL. The TTG codon found 7 bp downstream of a putative ribosome binding site (5′-AGGAGG-3′) with good complementarity to the 3′-terminal of *Bacillus subtilis* 16S rRNA[18] was regarded as the initiation codon of the AmyL gene. There were putative -35 and -10 regions, sequences of which are 5′-TTGTCA-3′ and 5′-TATAAT-3′, respectively,

Fig. 1

upstream of the ribosome binding site, and these regions were separated by 18 bp. An inverted-repeat sequence was found 6 bp downstream of the stop codon.

Twenty-eight amino acid residues at the N-terminal, which were not found in the native enzyme, are predicted to be a signal peptide by the Signal-P 4.0 program.[19] The deduced amino acid sequence of mature AmyL is comprised of 793 amino acid residues, and the molecular mass estimated from the sequence, 91,274 Da, coincided well with the molecular mass of 91 kDa for native AmyL.[5] The P1 sequence was also found in the deduced sequence of AmyL (Asp354-Lys357). A BLAST search of the mature AmyL sequence revealed that Gly395-Asp684 was assigned as the catalytic domain of α-amylase (PF00128),[20] and the four conserved sequences of GH family 13 enzymes were found in this region, as follows: region I, 464-YDFVPNH-470; region II, 542-GFRLDYAKG-550; region III, 571-FIFGEIWD-578; and region IV, 635-FLDNHD-640. Asp546, Glu575, and Asp640 are predicted to be the catalytic nucleophile, the general acid/base catalyst, and the second catalytic aspartic acid fulfilling multiple functions essential for catalysis, including elevating the pK_a of the general acid/base catalyst and promoting substrate distortion.[21] The catalytic domain of AmyL showed high sequence identity to neopullulanase-like α-amylases belonging to subfamily 20 of GH family 13,[4] although AmyL has no hydrolytic activity towards pullulan, unlike the other enzymes in this subfamily. The sequence identities of the catalytic domain of AmyL to *Geobacillus stearothermophilus* neopullulanase, *Thermoanaerobacter pseudethanolicus* cyclodextrinase, *Thermoactinomyces vulgaris* R-47 α-amylase I (TVAI), *T. vulgaris* R-47 α-amylase II (TVAII), and *Bacillus acidpullulyticus* maltogenic amylase were 37%, 37%, 33%, 36%, and 37%, respectively.

In contrast to neopullulanase-like α-amylases, the N-terminal region of AmyL (Gln29-Phe102) showed similarity to Carbohydrate-Binding Module family 20 (CBM20, PF00686), involved in binding to starch granules.[22] CBM20 is formed by a β-sandwich fold, and most CBM20s have two sugar-binding sites, site 1 and site 2. Two Trp residues, Trp45 and Trp84, of AmyL corresponded to the Trp residues conserved in

site 1 of CBM20, which formed a carbohydrate-binding platform (Fig. 2). Lys and Asn, forming direct hydrogen bonding interactions,[23] were also found at the equivalent positions of AmyL (Lys77 and Asn89). Trp61 of AmyL corresponded to the Trp residue conserved in site 2. Other amino acid residues responsible for binding to the carbohydrate in site 2 were not easily predicted, because this site is structurally more diverse than site 1. When the sequence of AmyL was compared with the starch-binding domain of *Aspergillus niger* glucoamylase (GA SBD), the best characterized enzyme in CBM family 20,[23] no other amino acid residues in site 2 that contribute to the interaction with the carbohydrate were found at the corresponding positions of AmyL.

Pro234-Thr301 and Tyr714-Gly780 were assigned to the chitobiose/β-hexosaminidase C-terminal domain (PF13290) and a domain of unknown function (DUF3459, PF11941), respectively, and their functions are not yet understood .

Fig. 2

Production and characterization of recombinant AmyL

Recombinant AmyL with a His-tag at the C-terminal (rAmyL) was produced in *E. coli* BL21 (DE3), and purified to homogeneity by Ni-chelating column chromatography. Sixteen mg of the purified enzyme, with a specific activity of 60.9 U/mg, was obtained from a 300-mL culture. The molecular mass of rAmyL was estimated to be 107 kDa, close to the value obtained by SDS-PAGE (91 kDa), by gel filtration column chromatography, indicating that rAmyL is a monomer in solution. rAmyL retained more than 90% of its maximum activity in a broad pH range, 7.8-10.5, as observed for the native enzyme (Fig. 3).[5] At high pH values, rAmy was less active than the native enzyme. It retained more than 90% of its original activity below 65°C (for 15 min at pH 9.5) and from pH 6.4 to 11.9 (for 24 h at 4ºC). At 70°C and higher temperatures, rAmy was more rapidly inactivated than the native enzyme. The native enzyme is stable in a pH range of 6.4-10.3, and thus rAmy was more stable at alkaline pH values.

The k_{cat} values of rAmyL for soluble starch, γ-CD, and maltotriose were 103 ± 9 s^{-1}, 67.6 ± 2.0 s^{-1}, and 5.33 ± 0.43 s^{-1}, respectively, and the K_m values were 0.100 ± 0.028

Fig. 3

mg/mL, 0.348 ± 0.019 mM, and 2.06 ± 0.33 mM, respectively.

To determine the predominant productive binding mode, the product distribution from the hydrolysis of pNPG5 was analyzed by measuring the releasing velocity of each aglycone having a *p*-nitrophenyl group. The releasing velocity of *p*-nitrophenyl α-glucoside was the most prominent at 53% of the entire reaction velocity (the sum of the reaction velocities of the aglycones), followed by *p*-nitrophenyl α-maltoside (34%). This result indicates that pNPG5 is hydrolyzed mainly through two productive binding modes covering subsites -4 to +2 and subsites -3 to +3.

A binding assay using raw cornstarch was carried out because a CBM20-like region was identified in the deduced amino acid sequence of AmyL. No decrease in enzyme activity of the supernatant was observed after incubation of 0.94-9.4 μM AmyL at 4°C for 1 h with gentle shaking in the presence of 100 mg/mL of starch granules (data not shown), and thus AmyL did not adsorb onto the starch granules.

Site-directed mutagenesis to three conserved Trp residues in the N-terminal CBM20-like sequence of AmyL

To confirm the function of the CBM-like region in the process of binding to soluble substrates, three conserved Trp residues (Trp45 and Trp84 from site 1 and Trp61 from site 2) were replaced with Ala. All the mutant enzymes were prepared as for wild-type AmyL, and the kinetic parameters for soluble starch, γ-CD, and maltotriose were measured (Table 1). The kinetic parameters for γ-CD were only minimally impacted by the mutations. Using maltotriose as substrate, W45A and W84A showed 1.7 to 1.9-fold lower K_m values, and W61A had a K_m similar to the wild type. In contrast to the reactions with oligosaccharides, W45A and W84A showed 2.2 to 2.4-fold lower k_{cat} values and 2.5-fold higher K_m values for soluble starch than the wild type. In contrast to the other two mutated AmyLs, W61A had a 0.8-fold and a 2.3-fold lower k_{cat} and K_m for soluble starch, respectively, resulting in a 1.9-fold higher k_{cat}/K_m than the wild type.

Table 1

Discussion

In this study, we analyzed the primary structure of AmyL, the enzymatic characteristics of recombinant AmyL, and the function of the N-terminal domain. The amino acid sequence of the central region of AmyL showed high similarity to the catalytic domains of neopullulanase-like α-amylases of GH family 13 subfamily 20 (Fig. 1). The calcium ion bound to the conserved binding site of typical α-amylases is not observed in neopullulanase-like α-amylases.[24] Similarly to neopullulanase-like α-amylases, AmyL also presumably does not have a calcium ion bound to the conserved binding site. The absence of a conserved calcium site in AmyL is thought to be the reason this enzyme showed high tolerance to chelating reagents. The amino acid residues responsible for thermostability and high resistance at alkaline pH were not easily predicted from the amino acid sequence of AmyL. Pro551 and Pro642, which were not found at the corresponding positions of the related enzymes, were predicted to be situated at the loops connecting the 4th β-strand and the 4th α-helix, and the 7th β-strand and the 7th α-helix, respectively. These residues might be important for high thermostability, because Pro residues situated at loops connecting secondary structures were postulated to be important for high thermostablity.[25]

In contrast to AmyL, which is a monomer in solution, many neopullulanase-like α-amylases are homodimers. The N-terminal domain of neopullulanase-like α-amylase is involved in the formation of the active site together with domain A of the other subunit.[26-28] The active-site clefts of dimeric neopullulanase-like α-amylases are narrow and deep, suitable for binding CDs.[26] The N-terminal domain of TVAI, also a monomer similar to AmyL, interacts strongly with domains A and B, unlike other neopullulanase-like α-amylases, to create the globular structure.[24] These structural differences are responsible for the variations in substrate specificity. TVAI has higher activities towards starch and pullulan, and lower activities towards α- and β-CDs than TVAII (a typical neopullulanase-like α-amylase) .[29,30] AmyL shows high activity

towards starch and γ-CD, and low activity towards α- and β-CDs,[5] similarly to TVAI. Although the N-terminal region of AmyL does not show any similarity to that of TVAI, AmyL is also a monomeric enzyme like TVAI, and the active-site cleft of AmyL might be shallow and more suitable for binding to starch, as was observed for TVAI.

TVAI has CBM34, which contains two maltooligosaccharide binding sites, at the N-terminal, in contrast to dimeric neopullulanase-like α-amylases.[31] It shows hydrolytic activity towards raw starch, and hence N-terminal CBM34 is considered to behave like a starch-binding domain. The N-terminal region of AmyL showed similarity to the portion of CBM20 involved in binding to starch granules. No adsorption of AmyL to starch granules was observed, but our site-directed mutational study revealed that the two Trp residues (Trp45 and Trp84) conserved in site 1 of CBM20 contribute to binding to a soluble long-chain substrate (Table 1), suggesting that the CBM20-like region of AmyL plays a role as substrate-binding domain. Both AmyL and TVAI have N-terminal sugar-binding domains followed by the catalytic domains, but AmyL has a substrate-binding domain structurally different from that of TVAI. To our knowledge, there is no report that enzymes belonging to GH family 13 subfamily20 have N-terminal CBM20, and hence AmyL is an enzyme having a novel domain organization. The W61A AmyL mutant showed higher catalytic efficiency for the hydrolysis of soluble starch than the wild type, and thus Trp61 of AmyL, corresponding to the conserved Trp residue in site 2 of CBM20, is not thought to be involved in sugar binding. Mutant AmyL showed a lower K_m value for soluble starch than the wild type, while the k_{cat} value of this mutant enzyme was similar to that of the wild type, suggesting that Trp61 disturbs the formation of the Michaelis complex. AmyL does not have a Tyr residue at the position corresponding to Tyr556 of GA SBD, which interacts mostly with carbohydrates,[23] and site 2 of AmyL presumably does not act as a sugar-binding site. In the complex structure of *Bacillus cereus* β-amylase and maltose, maltose was observed to be bound only at site 1 of CBM20 at the C-terminal of the enzyme,[32] and hence site 2 of CBM20 of this enzyme does not appear to possess binding affinity for carbohydrates.

Neopullulanase-like α-amylases hydrolyze pullulan, producing panose, but AmyL shows no hydrolytic activity towards this polysaccharide.[5] This is an essential difference in substrate specificity between AmyL and neopullulanase-like α-amylases. On comparison of the amino acid sequences of AmyL and neopullulanase-like α-amylases for which the three-dimensional structures have been solved, the loop connecting the 8th β-strand and the 8th α-helix in AmyL is longer than that in neopullulanase-like α-amylases, in which the conserved Asp on this loop is involved in the formation of subsite -2 (Fig. 4).[31] In neopullulanase-like α-amylases, short loop connecting the 8th β-strand and the 8th α-helix appears to make space for binding to pullulan, and thus this long loop of AmyL might result in steric hindrance that disturbs the process of binding to pullulan. Fig. 4

In this study, we found that AmyL has a catalytic domain similar to neopullulanase-like α-amylases, although it has no hydrolytic activity towards pullulan. In addition, we found that the N-terminal CBM20-like domain contributed to the process of binding to soluble long-chain substrates. Mutational analysis revealed that two Trp residues conserved at site 1 of CBM20 were important for high activity towards soluble long-chain substrates. Structural insight into the substrate binding mode and specificity should be pursued through the three-dimensional structure analysis of AmyL.

Acknowledgments

We thank Mr. Tomohiro Hirose of the Instrumental Analysis Division, Equipment Management Center, Creative Research Institution, Hokkaido University for amino acid analysis.

References

1) Ito S, Kobayashi T, Ara K, Ozaki K, Kawai S, and Hatada Y, *Extremophiles*, **2**,

185-190 (1998).
2) Cantarel B, Coutinho PM, Rancurel C, Bernard T, Lombard V, and Henrissat B, *Nucleic Acids Res.*, **37**, D233-D238 (2009).
3) MacGregor EA, Janecek S, and Svensson B, *Biochim. Biophys. Acta*, **1546**, 1-20 (2001).
4) Stam MR, Danchin EG, Rancurel C, Coutinho PM, and Henrissat B, *Protein Eng. Des. Sel.*, **19**, 555-562 (2006).
5) Kim DH, Morimoto N, Saburi W, Mukai A, Imoto K, Takehana T, Koike S, Mori H, and Matsui H, *Biosci. Biotechnol. Biochem.*, **76**, 1378-1383 (2012).
6) Horikoshi K, *Agric. Biol. Chem.*, **35**, 1783-1791 (1971).
7) Hayashi T, Akiba T, and Horikoshi K, *Agric. Biol. Chem.*, **52**, 443-448 (1988).
8) Boyer EW and Ingle MB, *J. Bacteriol.*, **110**, 992-1000 (1972).
9) Igarashi K, Hatada Y, Hagihara H, Saeki K, Takaiwa M, Uemura T, Ara K, Ozaki K, Kawai S, Kobayashi T, and Ito S, *Appl. Environ. Microbiol.*, **64**, 3282-3289 (1998).
10) Hagihara H, Igarashi K, Hayashi Y, Endo K, Ikawa-Kitayama K, Ozaki K, Kawai S, and Ito S, *Appl. Environ. Microbiol.*, **67**, 1744-1750 (2001).
11) Murakami S, Nishimoto H, Toyama Y, Shimamoto E, Takenaka S, Kaulpiboon J, Prousoontorn M, Limpaseni T, Pongsawasdi P, and Aoki K, *Biosci. Biotechnol. Biochem.*, **71**, 2393-2401 (2007).
12) Shirokizawa O, Akiba T, and Horikoshi K, *FEMS Microbiol. Lett.*, **58**, 131-135 (1990).
13) Pluskal MG, Przekop MB, Kavonian MR, Vecoli C, and Hicks DA, *BioTechniques*, **4**, 272-283 (1986).
14) Moore S and Stein WH, *J. Biol. Chem.*, **176**, 367-388 (1948).
15) Miller GL, *Anal. Chem.*, **31**, 426-428 (1959).
16) McFeeters RF, *Anal. Biochem.*, **103**, 302-306 (1980).
17) Miwa I, Okuda J, Maeda K, and Okuda G, *Clin. Chim. Acta*, **37**, 538-540 (1972).
18) Murray CL and Rabinowitz JC, *J. Biol. Chem.*, **257**, 1053-1062 (1982).

19) Petersen TN, Brunak S, von Heijne G, and Nielsen H, *Nat. Methods*, **8**, 785-786 (2011).
20) Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, and Finn RD, *Ncleic Acids Res.*, **40**, D290-301 (2012).
21) Uitdehaag JCM, Mosi R, Kalk KH, van der Veen BA, Dikhuizen L, Withers SG, and Dijkstra BW, *Nat. Struct. Biol.*, **6**, 432-436 (1999).
22) Christiansen C, Abou Hachem M, Janecek S, Viksø-Nielsen A, Blennow A, and Svensson B, *FEBS J.*, **276**, 5006-5029 (2009).
23) Sormachi K, Le Gal-Coëffet MF, Williamson G, Archer DB, and Williamson MP, *Structure*, **5**, 647-661 (1997).
24) Kamitori S, Abe A, Ohtaki A, Kaji A, Tonozuka T, and Sakano Y, *J. Mol. Biol.*, **318**, 443-453 (2002).
25) Watanabe K, Chishiro K, Kitamura K, and Suzuki Y, *J. Biol. Chem.*, **266**, 24287-24294 (1991).
26) Kim JS, Cha SS, Kim HJ, Kim TJ, Ha NC, Oh ST, Cho HS, Cho MJ, Kim MJ, Lee HS, Kim JW, Choi KY, Park KH, and Oh BH, *J. Biol. Chem.*, **274**, 26279-26286 (1999).
27) Kamitori S, Kondo S, Okuyama K, Yokota T, Shimura Y, Tonozuka T, and Sakano Y, *J. Mol. Biol.*, **287**, 907-921 (1999).
28) Hondoh H, Kuriki T, and Matsuura Y, *J. Mol. Biol.*, **326**, 177-188 (2003).
29) Ibuka A, Tonozuka T, Matsuzawa H, and Sakai H, *J. Biochem.*, **123**, 275-282 (1998).
30) Shimura Y, Tonozuka T, and Sakano Y, *J. Appl. Glycosci.*, **45**, 393-395 (1998).
31) Abe A, Tonozuka T, Sakano Y, and Kamitori S, *J. Mol. Biol.*, **335**, 811-822 (2004).
32) Mikami B, Adachi M, Kage T, Sarikaya E, Nanmori T, Shinke R, and Utsumi S, *Biochemistry*, **38**, 7050-7061 (1999).

Figure legends

Fig. 1. DNA and Amino Acid Sequences of AmyL.

Numbers of amino acids and nucleotides are shown to the left and right, respectively. Asterisk indicates the stop codon. The N-terminal and partial internal amino acid sequences of AmyL purified from the culture supernatant of *Bacillus* sp. AAH-31 are shown in bold face. The ribosome binding site (RBS), inverted repeat, and -35 and -10 regions are underlined. *Hin*dIII and *Bam*HI sites are boxed. Conserved regions of GH family 13 enzymes are indicated by double underlining. Arrows indicate primers used to obtain *AmyL*.

Fig. 2. Comparison of the N-Terminal Sequence of AmyL with CBM20.

The amino acid sequences of the N-terminal region of AmyL and CBM20s, the three-dimensional structures of which are known, were aligned with the Clustal W program (http://clustalw.ddbj.nig.ac.jp/). Amino acid residues conserved in sites 1 and 2 are shown in boldface, and the circle and triangle indicate site 1 and site 2, respectively. AnGA, *A. niger* glucoamylase (Genbank ID: CAK38411.1); BcBA, *B. cereus* β-amylase (BAA34650.1); Bc251CGT, *B. circulans* 251 cyclodextrin glucanotransferase (CGTase, CAA55023.1); Bc8CGT, *B. circulans* 8 CGTase (CAA48401.1); GsCGT, *G. stearothermophilus* NO. 2 CGTase (CAA41770.1); TtCGT, *Thermoanaerobacterium thermosulfurigenes* EM1 CGTase (AAB00845.1); Novamyl, *G. stearothermophilus* C599 maltogenic α-amylase (ABA38666.1); G4-Amy, *Pseudomonas stutzeri* MO-19 maltotetraose-forming α-amylase (AAA25707.1).

Fig. 3. Comparison of Physicochemical Properties between Native and Recombinant AmyL.

The pH activity curve (a), pH stability (b), and thermal stability (c) of the native (●) and recombinant (○) AmyL were compared. Data for the native enzyme are from

reference 5.

Fig. 4. Comparison of Partial Amino Acid Sequences between β-Strand 8 and α-Helix 8 of AmyL and Related Enzymes.

The amino acid sequences of AmyL and related enzymes were aligned with the Clustal W program. Amino acid residues completely conserved in the enzymes are shown in bold face. TVAI, *T. vulgaris* R-47 α-amylase I (Genbank ID: BAA02471.1); TVAII, *T. vulgaris* α-amylase II (BAA02473.1); NPU, *G. stearothermophilus* neopullulanase (AAA22622.1); TMA, *Thermus* sp. IM6501 (AAC15072.1); BCD, *Bacillus* sp. cyclomaltodextrinase (AAA92925.1).

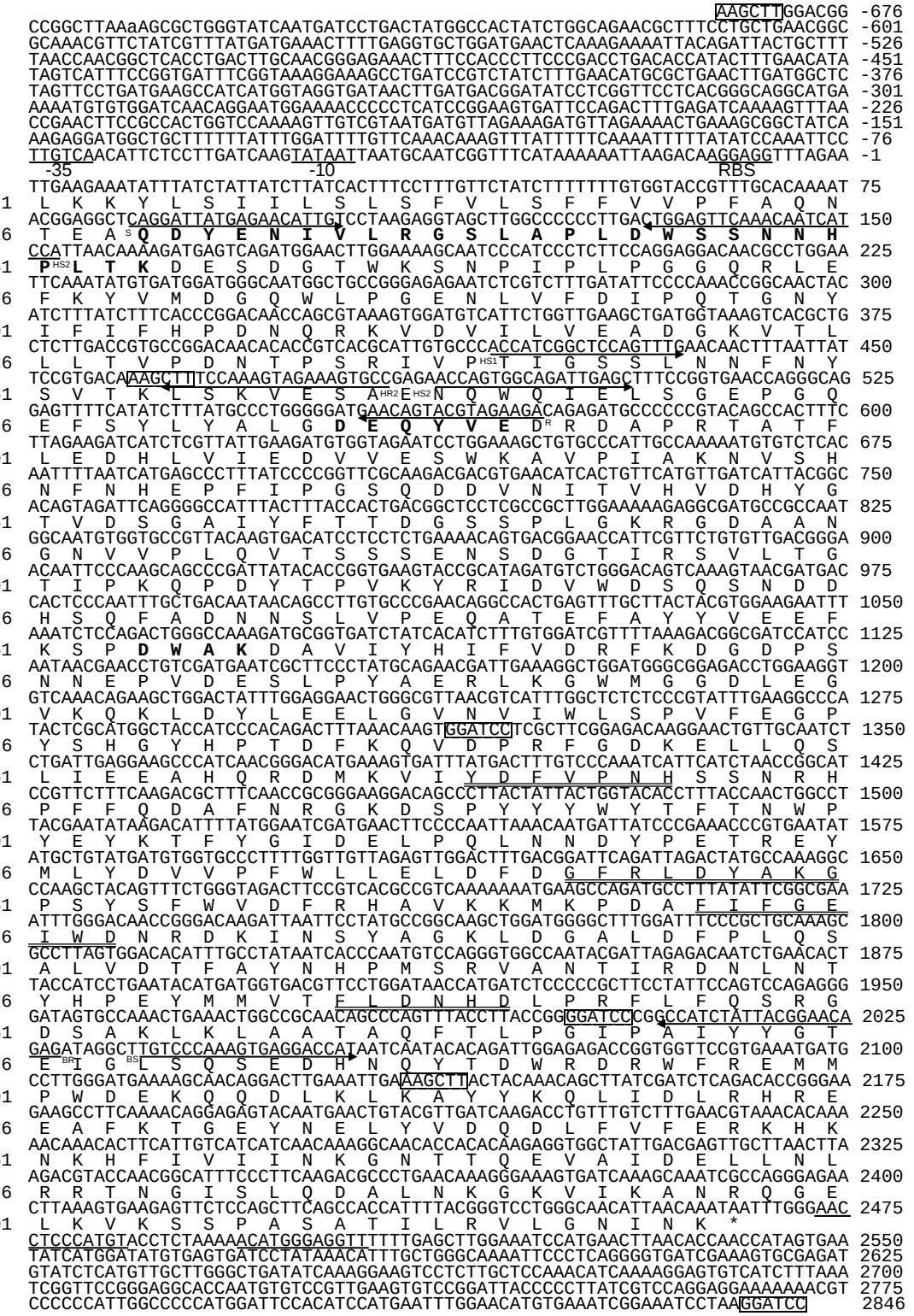

Fig. 1, Saburi et al.

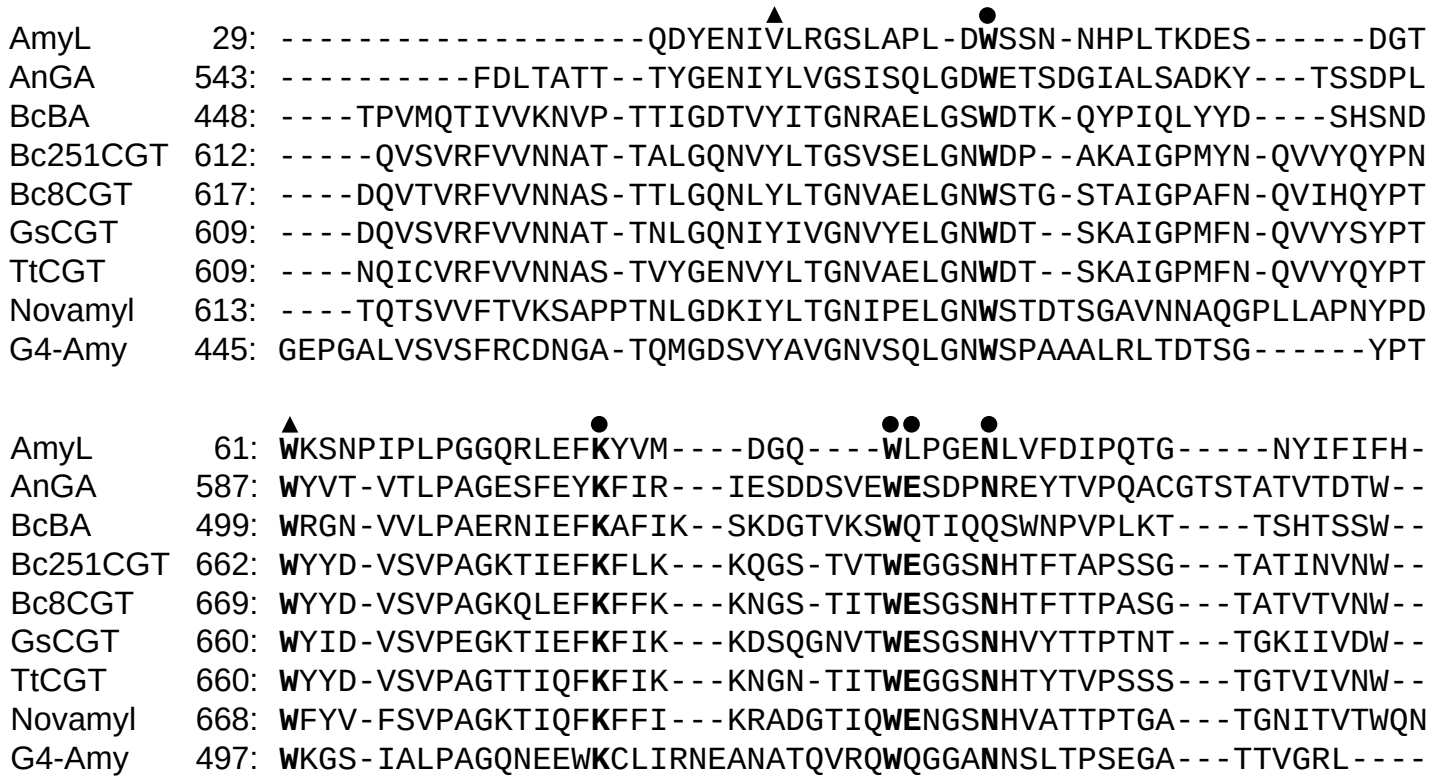

Fig. 2, Saburi et al.

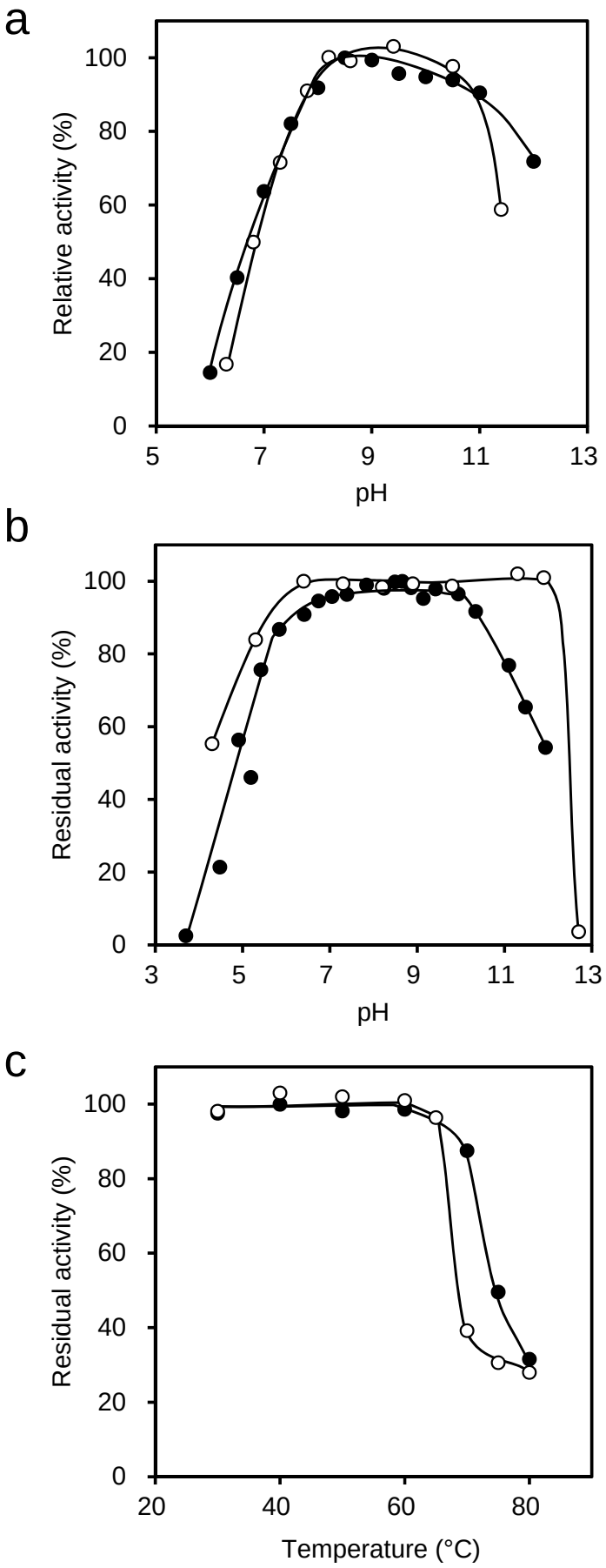

Fig. 3., Saburi, et al.

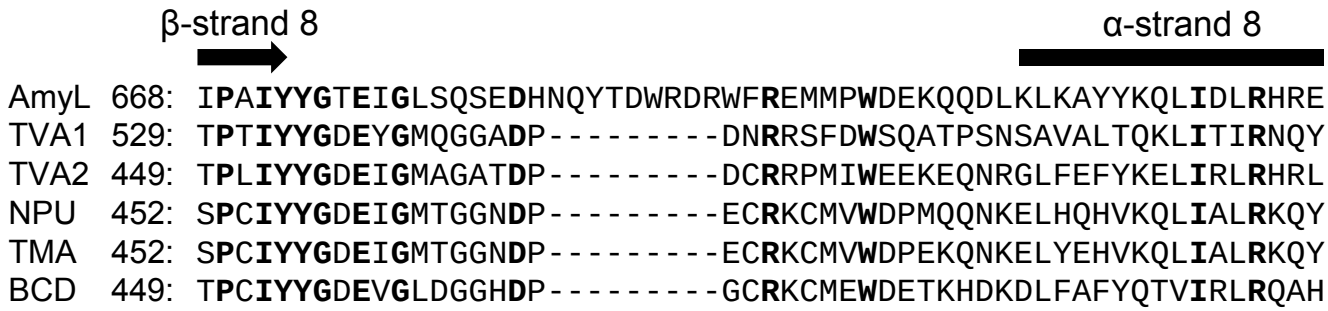

Fig. 4, Saburi et al.

Table 1. Comparison of Kinetic Parameters between Trp Mutated AmyL and the Wild Type

	Soluble starch			γ-CD			Maltotriose		
Enzyme	k_{cat} (s^{-1})	K_m ($mg \cdot mL^{-1}$)	k_{cat}/K_m ($s^{-1} \cdot mg^{-1} \cdot mL$)	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($s^{-1} \cdot mM^{-1}$)	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($s^{-1} \cdot mM^{-1}$)
Wild type	103 ± 9	0.100 ± 0.028	1030	67.6 ± 2.0	0.348 ± 0.019	194	5.33 ± 0.43	2.06 ± 0.33	2.59
W45A	42.9 ± 4.1	0.251 ± 0.028	171	76.9 ± 6.0	0.503 ± 0.032	153	4.26 ± 0.20	1.21 ± 0.18	3.52
W61A	84.6 ± 7.0	0.0439 ± 0.0151	1930	69.5 ± 2.2	0.353 ± 0.039	197	4.71 ± 0.32	1.82 ± 0.44	2.59
W84A	47.7 ± 6.8	0.248 ± 0.026	192	73.0 ± 7.3	0.381 ± 0.038	192	4.05 ± 0.10	1.10 ± 0.09	3.68

Data are mean ± standard deviation for three independent experiments.