Regular paper

Substrate specificity of glycoside hydrolase family 1 β-glucosidase AtBGlu42 from *Arabidopsis thaliana* and its molecular mechanism

Shu Horikoshi[1], Wataru Saburi[1], Jian Yu[2], Hideyuki Matsuura[1], James R. Ketudat Cairns[3], Min Yao[2], and Haruhide Mori[1] *

[1]Research Faculty of Agriculture, Hokkaido University, Kita 9 Nishi 9, Sapporo 060-8589, Japan.
[2]Faculty of Advanced Life Science, Hokkaido University, Kita 10 Nishi 8, Sapporo 060-0810, Japan.
[3]School of Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand.

Correspondence: Haruhide Mori, hmori@chem.agr.hokudai.ac.jp

Running head: Enzymatic and structural analysis of AtBGlu42

Abstract

Plants possess many glycoside hydrolase family 1 (GH1) β-glucosidases, which physiologically function in cell wall metabolism and activation of bioactive substances, but most remain uncharacterized. One GH1 isoenzyme AtBGlu42 in *Arabidopsis thaliana* has been identified to hydrolyze scopolin using the gene deficient plants, but no enzymatic properties were obtained. Its sequence similarity to another functionally characterized enzyme Os1BGlu4 in rice suggests that AtBGlu42 acts on also oligosaccharides. Here, we show that the recombinant AtBGlu42 possesses high k_{cat}/K_m not only on scopolin, but also on various β-glucosides, cellooligosaccharides, and laminarioligosaccharides. Of the cellooligosaccharides, cellotriose is the most preferred. The crystal structure, determined at 1.7 Å resolution, suggests that Arg342 gives unfavorable binding to cellooligosaccharides at subsite +3. The mutants R342Y and R342A showed the highest preference on cellotetraose or cellopentaose with increased affinities at subsite +3, indicating that the residues at this position have an important role for chain length specificity.

Keywords: β-Glucosidase, Glycoside hydrolase family 1, Substrate specificity, X-ray crystallography, *Arabidopsis thaliana*

35 **Introduction**

36 β-Glucosidase (EC 3.2.1.21) catalyzes the hydrolysis of β-glucosidic linkages at the non-
37 reducing end of substrates to release D-glucose. This enzyme is ubiquitous in archaea, eubacteria, and
38 eukaryotes, and contributes to the degradation of cell wall, metabolism of glycolipids, and activation
39 of bioactive substances (Ketudat Cairns and Esen 2010; Ketudat Cairns *et al.* 2015). Based on the
40 sequence-based classification of glycoside hydrolases, β-glucosidases are in the glycoside hydrolase
41 families (GHs): GH1, GH2, GH3, GH5, GH30, GH39, and GH116 (Lombard *et al.* 2014). Most of
42 these families, excluding GH3 and GH116, fall into GH Clan A. Clan A enzymes share a $(\beta/\alpha)_8$-barrel
43 catalytic domain. Their catalytic residues acting as general acid/base and nucleophile are Glu residues
44 at the C-terminal of β-strands 4 and 7, respectively (Jenkins *et al.* 1995; Henrissat *et al.* 1995; Rye and
45 Withers 2000).

46 The substrate specificity of GH1 β-glycosidases is very diverse. Most GH1 enzymes act on β-
47 glucosides, but β-D-fucosides, β-D-galactosides, β-D-xylosides, and β-D-mannosides are also their
48 substrates (Ketudat Cairns and Esen 2010). GH1 enzymes also exhibit a variety of specificities for the
49 aglycone of β-glycosides, and oligosaccharides with different linkages and chain lengths (Opassiri *et*
50 *al.* 2004; Seshadri *et al.* 2009; Opassiri *et al.* 2010; Rouyi *et al.* 2014). Plant GH1 β-glucosidases act
51 on various β-glucosides of secondary metabolites, including plant hormones such as salicylic acid
52 (Himeno *et al.* 2013), tuberonic acid (Wakuta *et al.* 2010; Wakuta *et al.* 2011), abscisic acid (Lee *et al.*
53 2006), and gibberellin (Hua *et al.* 2013), benzoxazinoids (Babcock and Esen 1994; Sue, Ishihara and
54 Iwamura 2000), cyanohydrins (Hösel *et al.* 1987), alkaloids (Barleben *et al.* 2007; Xia *et al.* 2012),
55 and phenylpropanoids (Ahn *et al.* 2010; Roepke and Bozzo 2015; Baba, Vishwakarma and Ashraf
56 2017). The enzymes are probably involved in quantitative regulation of the active forms of those
57 compounds through deglycosylation (Ketudat Cairns *et al.* 2015). Plants possess so many GH1
58 isoenzymes compared with other organisms (Ketudat Cairns *et al.* 2012). Arabidopsis and rice have 40
59 and 34 functional genes encoding GH1 enzymes, respectively (Xu *et al.* 2004; Opassiri *et al.* 2006).
60 This multiplicity is possibly associated with the regulation of the various compounds through their

61 specific expression and substrate specificities of the isoenzymes, although most of their physiological
62 functions remain to be clarified.
63 A GH1 β-glucosidase from Arabidopsis, AtBGlu42, is likely localized in the cytoplasm because
64 it has no N-terminal signal peptide in its precursor sequence, while the other Arabidopsis GH1
65 isoenzymes are predicted to possess one (Xu *et al.* 2004). AtBGlu42 is known to be involved in the
66 acquisition of rhizobacteria-induced systemic resistance and iron-uptake through deglycosylation of
67 phenolic compounds, mainly the coumarin β-glucoside, scopolin. These functions were predicted on
68 the basis of expression of the gene and the phenotypes of *AtBGlu42*-modified plants (Zamioudis,
69 Hanson and Pieterse 2014; Stringlis *et al.* 2018). On the other hand, AtBGlu42 shares high sequence
70 identity (64%) with the rice cytosolic GH1 β-glucosidase, Os1BGlu4 (Rouyi *et al.* 2014), which has
71 high hydrolytic activity towards laminaribiose (β1-3-linked) and β1-4-linked cellooligosaccharides of
72 degree of polymerization (DP) 3-4 in addition to β-glucosides such as salicin and esculin. These imply
73 that AtBGlu42 may act on not only scopolin but also various β-glucosides including oligosaccharides.
74 In this study, we investigated enzymatic characteristics and crystal structure of AtBGlu42 using
75 recombinant enzyme produced in *Escherichia coli*. An amino acid residue discriminating the enzyme
76 from other GH1 β-glucosidases in terms of cellooligosaccharide preference was identified in the three-
77 dimensional structure and verified through the site-directed mutagenesis. The local domain structure
78 determining the spatial location of the residues is also discussed.
79
80 **Materials and methods**
81 **Plant materials**
82 *A. thaliana* ecotype Columbia was grown on half-strength Murashige and Skoog medium (pH
83 5.7; Wako Pure Chemical Industries, Osaka, Japan) with 10 g L^{-1} sucrose and 3 g L^{-1} gellan gum
84 (Wako Pure Chemical Industries) for 10 days at 23 °C under 16 h light/8 h dark conditions.
85
86 **AtBGlu42 expression plasmid**

87 Total RNA was prepared from the seedlings of *A. thaliana* with RNeasy Mini Kit (Qiagen,
88 Hilden, Germany), and cDNA was synthesized with Superscript III First-Strand Synthesis System for
89 RT-PCR (Thermo Fisher Scientific, Waltham, MA, USA). The AtBGlu42 cDNA was amplified by
90 PCR with the primers, 5′-GCTTCTTAAGTCTGTCTCTCTCTTC-3′ (sense) and 5′-
91 ACGAAACACITAGTCAAAATATGAG-3′ (antisense), and PrimeSTAR HS DNA polymerase
92 (Takara Bio, Kusatsu, Japan), followed by reamplification with the primers, 5′-
93 GTGCCACGCGGTTCTATGGCACAGAAGCTTAACTT-3′ (sense, the 15-bp overlap with pET-32a
94 underlined) and 5′-CGCAAGCTTGTCGACTCATTCCTTCTTACCTTTGT-3′ (antisense). The
95 amplified DNA was inserted into the pET-32a vector (Novagen, Darmstadt, Germany) using In-Fusion
96 HD Cloning Kit (Takara Bio). The linear pET-32a was prepared by PCR using the primers, 5′-
97 GTCGACAAGCTTGCGGCCGC-3′ (sense) and 5′-AGAACCGCGTGGCACCAGAC-3′ (antisense).
98 Expression plasmids of AtBGlu42 mutants were prepared using a PrimeSTAR Mutagenesis Basal Kit
99 (Takara Bio) with the following primers: for R342Y, 5′-TTGGAGTATATTGTTGAACTGGAAAAT-3′
100 (sense, substituted nucleotides underlined) and 5′-AACAATATACTCCAATTCTTGTGCTTG-3′
101 (antisense); and for R342A, 5′-TTGGAGGCGATTGTTGAACTGGAAAAT-3′ (sense) and 5′-
102 AACAATCGCCTCCAATTCTTGTGCTTG-3′ (antisense). The DNA sequences encoding AtBGlu42
103 and neighboring regions were verified in all the expression constructs by DNA sequencing using an
104 Applied Biosystems 3130 Genetic Analyzer (Thermo Fisher Scientific). The AtBGlu42-coding
105 sequence was identical to the deposited cDNA sequence (Locus tag: AT5G36890.1).
106
107 **Preparation of recombinant AtBGlu42**
108 Recombinant AtBGlu42 was produced in *E. coli* as a fusion protein with thioredoxin and a His_6
109 tag at its N-terminus. *E. coli* Origami B(DE3) transformant, harboring the AtBGlu42 expression
110 plasmid, was cultured in 1.0 L of Luria-Bertani medium containing 100 μg mL^{-1} ampicillin at 37 °C
111 until the culture A_{600} reached 0.5. Recombinant protein production was induced by the addition of
112 isopropyl β-D-1-thiogalactpyranoside (Wako Pure Chemical Industries) to be 0.1 mM in the medium
113 and the incubation continued at 20 °C for 20 h with vigorous shaking. The *E. coli* cells were harvested

114 by centrifugation (7,000 × *g*, 4 °C, 10 min), and suspended in 30 mL of 20 mM sodium phosphate
115 buffer (pH 7.5) containing 0.5 M NaCl (buffer A). The bacterial cells were disrupted by sonication
116 using a Sonifier 450 (Branson, Danbury, CT, USA), and the supernatant was obtained by
117 centrifugation (32,000 × *g*, 4 °C, 10 min). It was loaded onto a Ni immobilized Chelating Sepharose
118 Fast Flow column (1.6 cm i.d. × 2.5 cm, 5 mL; GE Healthcare, Uppsala, Sweden), equilibrated with
119 buffer A. After thoroughly washing the column with buffer A containing 30 mM imidazole, the
120 adsorbed protein was eluted by a linear gradient of imidazole from 30 to 500 mM in buffer A (total
121 elution volume, 200 mL). Fractions containing highly purified enzyme were pooled. Purity of the
122 protein was confirmed by SDS-PAGE. The collected sample was dialyzed against 20 mM sodium
123 phosphate buffer (pH 7.0) and concentrated to 2.49 mg mL^{-1} using an Amicon Ultra-15 centrifugal
124 filter (30,000 nominal molecular weight limits; Merck Millipore, Billerica, MA, USA). The
125 preparation was stored at 4 °C until use. The AtBGlu42 mutants were prepared in the same fashion as
126 the wild type.
127 For X-ray crystallography, recombinant AtBGlu42 obtained as above, but from 4 L
128 fermentation, was digested with thrombin (Sigma Aldrich, St. Louis, MO, USA) to remove its N-
129 terminal region containing thioredoxin and a His_6 tag. Thrombin was added to the enzyme at a protein
130 ratio of 1:20 (w/w) and it was incubated for 20 h at 4 °C. The untagged AtBGlu42 was collected in the
131 non-adsorbed fractions in the Ni-affinity column chromatography done as described above, dialyzed
132 against 20 mM Tris-HCl buffer (pH 7.2) (buffer B), and further purified by anion exchange column
133 chromatography with a DEAE Sepharose Fast Flow column (2.5 cm i.d. × 16 cm, 78 mL; GE
134 Healthcare) equilibrated with buffer B. After washing with buffer B, adsorbed protein was eluted with
135 a linear gradient of NaCl from 0 to 0.5 M in buffer B (400 mL). The highly purified fractions were
136 pooled, dialyzed against buffer B, and concentrated to 16.1 mg mL^{-1} as described above. The purified
137 enzyme preparations were stored at 4 °C until use.
138 The protein concentration of the purified enzymes was determined based on the molar
139 quantities of each amino acid measured using the ninhydrin colorimetric method with JLC-500/V
140 (JEOL, Tokyo, Japan) after complete acid hydrolysis of the enzyme in 6 M HCl at 110 °C for 24 h.

141

142 **Standard enzyme assay**

143 Enzyme activity was measured using *p*-nitrophenyl β-D-glucopyranoside (pNP β-Glc, Sigma
144 Aldrich) as substrate. A reaction mixture (50 μL), consisting of an appropriate concentration of
145 enzyme, 50 mM sodium phosphate buffer (pH 6.8), 0.2 mg mL^{-1} bovine serum albumin (BSA), and 1
146 mM pNP β-Glc, was incubated at 30 °C for 10 min. The reaction was terminated by adding 25 μL of 2
147 M Na_2CO_3, and A_{400} was measured to determine the *p*-nitrophenol (pNP) with molar extinction
148 coefficient of 18750 M^{-1} cm^{-1}. One unit (U) of β-glucosidase activity was defined as the amount of
149 enzyme that releases 1 μmol of pNP in 1 min under these conditions.

150

151 **Evaluation of effects of pH and temperature on activity and stability**

152 Optimal pH was determined as the standard activity assay but using 80 mM modified Britton-
153 Robinson buffer (pH 4.1-11.0; mixture of 80 mM sodium acetate, sodium phosphate, and glycine-
154 NaOH buffers) in place of 50 mM sodium phosphate buffer. Optimum temperature was determined by
155 measuring activities at 25-60 °C. The pH stability was determined based on residual activity after
156 keeping 0.89 μM enzyme in 195 mM modified Britton-Robinson buffer (pH 4.1-11.2) at 4 °C for 24 h.
157 Temperature stability was determined based on residual activity after keeping 24.2 nM enzyme in 12.5
158 mM sodium phosphate buffer (pH 7.0) containing 1 mg mL^{-1} BSA at 4-60 °C for 15 min. The ranges
159 of pH and temperature in which the enzyme retained more than 90% of its original activity were
160 defined the stable ranges.

161

162 **Substrate specificity**

163 Reaction rates to the following substrates (2 mM) were measured under the conditions of the
164 standard enzyme assay: pNP β-Glc, pNP β-D-fucopyranoside (pNP β-D-Fuc), pNP β-D-
165 galactopyranoside (pNP β-Gal), pNP β-D-mannopyranoside (pNP β-Man), pNP β-D-xylopyranoside
166 (pNP β-Xyl), cellobiose, sophorose, pNP β-cellobioside, arbutin, phlorizin (Sigma Aldrich),
167 cellotriose, cellotetraose, cellopentaose, cellohexaose, laminaribiose, laminaritriose, laminaritetraose

168 (Megazyme, Bray, Ireland), gentiobiose, methyl β-D-glucopyranoside (Nacalai Tesque, Kyoto, Japan),
169 neotrehalose (Hayashibara, Okayama, Japan), 4-methylumbelliferyl β-D-glucopyranoside (4MU β-
170 Glc), helicin, octyl β-D-glucopyranoside (Tokyo Chemical Industry, Tokyo, Japan), scopolin (Indofine
171 Chemical Company, Hillsborough, NJ, USA), phenyl β-D-glucopyranoside (Kanto Chemical, Tokyo,
172 Japan), and salicylic acid β-D-glucopyranoside (Grynkiewicz *et al.* 1993). pNP release rates from pNP
173 glycosides, except pNP β-cellobioside, and D-glucose release rates from the others were measured. D-
174 Glucose was quantified using the Glucose CII Test (Wako Pure Chemical Industries) after the enzyme
175 reactions (50 μL) were terminated by the addition of 25 μL of 4 M Tris-HCl buffer (pH 7.0).
176 Kinetic parameters of the Michaelis-Menten equation were determined using non-linear
177 regression in the s-*v* plots, in which substrate concentrations were as follows: 0.05-0.6 mM for pNP β-
178 Glc; 0.05-0.8 mM for pNP β-D-Fuc, 4MU β-Glc, cellotriose, cellotetraose, cellopentaose,
179 cellohexaose, laminaritriose, and laminaritetraose; 0.5-8 mM for pNP β-Gal; 0.1-1.6 mM for pNP β-
180 Man; 0.125-2 mM for pNP β-Xyl, laminaribiose, and sophorose; 0.75-12 mM for cellobiose; 1-12 mM
181 for gentiobiose; 0.0125-0.4 mM for pNP β-cellobioside and helicin; and 0.025-0.4 mM for scopolin.
182 Fitting was done using Grafit version 7.0.2 (Erithacus Software, East Grinstead, UK). Subsite
183 affinities for binding to cellooligosaccharides were calculated from the kinetic parameters according
184 to the subsite theory for exo-glycosidase (Hiromi *et al.* 1973; Saburi *et al.* 2006).
185
186 **Kinetic analysis of the reaction with pNP β-Glc**
187 The reaction for pNP β-Glc was analyzed by nonlinear regression based on the kinetic model of
188 retaining glycosidases that catalyze both hydrolysis and transglycosylation (Kobayashi *et al.* 2011;
189 Saburi *et al.* 2013). Reaction equations used are as follows:
190 $v_{ag} = v_h + v_{tg}$ (eq. 1)
191 $v_h = k_{cat1} K_{m2} s / (s^2 + K_{m2} s + K_{m1} K_{m2})$ (eq. 2)
192 $v_{tg} = k_{cat2} s^2 / (s^2 + K_{m2} s + K_{m1} K_{m2})$ (eq. 3)
193 $r_{tg} = v_{tg} / v_{ag} = s / (K_{TG} + s)$ (eq. 4)

194 Here, s is pNP β-Glc concentration; v_{ag}, v_h, and v_{tg} are reaction rates of aglycone release, hydrolysis,
195 and transglucosylation, respectively; r_{tg} is transglucosylation ratio, and K_{TG} is transglycosylation
196 parameter. The kinetic parameters k_{cat1}, k_{cat2}, K_{m1}, and K_{m2} are defined as follows by the rate constants
197 shown in Figure 1:

198 $k_{cat1} = k_1k_2k_3(k_{-4} + k_5)/\{k_4k_5(k_{-1} + k_2) + k_1(k_{-4} + k_5)(k_2 + k_3)\}$

199 $k_{cat2} = k_2k_5/(k_2 + k_5)$

200 $K_{m1} = k_3(k_{-1} + k_2)(k_{-4} + k_5)/\{k_4k_5(k_{-1} + k_2) + k_1(k_{-4} + k_5)(k_2 + k_3)\}$

201 $K_{m2} = \{k_4k_5(k_{-1} + k_2) + k_1(k_{-4} + k_5)(k_2 + k_3)\}/\{k_1k_4(k_2 + k_5)\}$

202 A reaction mixture (250 μL), containing 14.3 nM AtBGlu42, 50 mM sodium phosphate buffer
203 (pH 6.8), 0.2 mg mL^{-1} BSA, and 0.125-8 mM pNP β-Glc, was incubated at 30 °C for 10 min. Aliquots
204 (100 μL each) were taken, and the reaction was stopped with 200 μL of 1 M Na_2CO_3 and 50 μL of 4 M
205 Tris-HCl buffer (pH 7.0) to measure pNP and D-glucose, respectively, as described above. The
206 parameters, k_{cat1}, K_{m1}, and K_{m2}, were determined by nonlinear regression of equation 2 to the data.
207 Then, k_{cat2} was determined from equations 1 and 3.
208

209 **TLC analysis of reaction products from pNP β-Glc**

210 A reaction mixture (200 μL), containing 2.27 μM AtBGlu42, 50 mM sodium phosphate buffer
211 (pH 6.8), 0.2 mg mL^{-1} BSA, and 8 mM pNP β-Glc, was incubated at 30 °C for 30 min. Aliquots (10
212 μL) were taken at the indicated times and incubated at 100 °C for 2 min to terminate the reaction. TLC
213 was done on Silica-gel (Aluminum TLC plate, Silica gel 60 F_{254}, Merck, Darmstadt, Germany) using a
214 developing solvent chloroform/methanol/water (14/6/1; v/v/v). Spots of pNP glycosides were
215 visualized by UV irradiation and carbohydrates were detected by spraying a detection reagent acetic
216 acid/sulfuric acid/anisaldehyde (100/2/1; v/v/v) and heating.
217

218 **Crystallization and data collection**

219 Crystallization of the non-tagged AtBGlu42 was performed by the sitting-drop vapor diffusion
220 method, in which 0.75 μL of protein solution (16.1 mg mL^{-1} in 20 mM Tris-HCl buffer, pH 7.2) was

221 mixed with an equal volume of reservoir solution containing 0.1 M sodium cacodylate buffer (pH 6.5),
222 0.2 M calcium acetate, and 18% (w/v) polyethylene glycol 8000 (Sigma Aldrich), and incubated at
223 20 °C for 2 months. X-ray diffraction data of AtBGlu42 were collected on the beam-line BL41XU at
224 SPring-8 (Sayo, Japan). The data sets were indexed, integrated, scaled, and merged using the XDS
225 program suite (Kabsch 2010). The asymmetric unit of AtBGlu42 contained one molecule
226 corresponding to a Matthews coefficient (Matthews 1968) of 2.20 $Å^3 Da^{-1}$ and an estimated solvent
227 content of 44.2%. All the data collection statistics are summarized in **Table 1**.
228
229 **Structure solution and refinement**
230 The structure of AtBGlu42 was determined by the molecular replacement method with the
231 program Phaser in the PHENIX program package (McCoy *et al.* 2007; Adams *et al.* 2010). The
232 structure of Os3BGlu6 (52% sequence identity to AtBGlu42; PDB entry, 3GNO) (Seshadri *et al.*
233 2009) was used as the search model. Several rounds of refinement were performed using the program
234 PHENIX.REFINE in the PHENIX program package, alternating with manual fitting and rebuilding
235 based on $2F_o–F_c$ and $F_o–F_c$ electron densities in COOT (Emsley and Cowtan 2004). Water molecules
236 and glycerol were built based on $2F_o–F_c$ and $F_o–F_c$ electron densities. The final refinement statistics
237 and geometry defined by MOLPROBITY (Chen *et al.* 2010) are shown in **Table 1**. The atomic
238 coordinates and structure factors were deposited in the Protein Data Bank (PDB entry, 7F3A). All
239 structure figures were generated by PYMOL ver. 2.1 (Schrödinger, LLC, New York, NY, USA).
240
241 **Molecular docking**
242 Molecular docking was performed using AutoDock Vina (ADT version 1.5.6) (Trott and Olson
243 2010) to predict the interaction between AtBGlu42 and scopolin. Scopolin was prepared with the
244 glucosyl moiety in the $^1S_3$ skew boat conformation with the Discovery Studio 4.0 program (Dassault
245 Systèmes BIOVIA, San Diego, CA, USA). Water molecules and glycerol molecules in the active site
246 of AtBGlu42 were removed for the docking test. The scopolin was docked into the active site of

247 AtBGlu42 using the grid box with a spacing of 1 Å and dimension of 40×40×40, centered at positions
248 of 15.80 (x), 63.04 (y), 43.12 (z). The exhaustiveness value was set to 8.
249
250 **Results**
251 **Preparation of recombinant AtBGlu42**
252 A cDNA prepared from *A. thaliana* seedling RNA was used to produce AtBGlu42 in
253 recombinant *E. coli* as a fusion protein with a 131-residue-long vector-derived region containing
254 thioredoxin and a His_6 tag on the N-terminal side of Met1 of AtBGlu42. From the cell-free extract
255 obtained from 1 L of culture, 8.8 mg of the recombinant enzyme was purified by immobilized
256 metal (Ni) affinity column chromatography. The molecular mass of recombinant AtBGlu42,
257 estimated by SDS-PAGE, was 70 kDa, and it coincided with the theorical mass from the amino
258 acid sequence (70,143 Da) (**Figure 2**). Its specific activity was 10.5 U mg^{-1}. The optimum pH and
259 temperature were pH 6.8 and 40 °C, respectively (**Figure 3a and b**). AtBGlu42 retained over 90%
260 of the original activity after incubation at pH 6.3-9.3 at 4 °C for 24 h, or up to 35 °C at pH 7.0 for
261 15 min (**Figure 3c and d**).
262
263 **Specificity to pNP β-glycosides**
264 The glycone specificity of AtBGlu42 was evaluated based on the reaction velocities to 2 mM
265 pNP β-glycosides and kinetic parameters (**Table 2**). pNP β-Glc and pNP β-D-Fuc were good
266 substrates, but AtBGlu42 acted also on β-galactoside, β-mannoside, and β-xyloside with lower
267 velocities. AtBGlu42 catalyzed transglucosylation for pNP β-Glc. In the early stage of the reaction
268 with 8 mM pNP β-Glc, transglucosylation products, including pNP β-cellobioside, were observed in
269 the TLC analysis, and the substrate and transglucosylation products were completely hydrolyzed
270 within 30 min of reaction (**Figure 4a and b**). Velocities for aglycone release (v_{ag}), hydrolysis (v_h), and
271 transglucosylation (v_{tg}) were calculated from velocities for the liberation of pNP and D-glucose
272 (**Figure 4c**). These reaction velocities followed the rate equations 1-3 of a retaining glycoside
273 hydrolase catalyzing both hydrolysis and transglucosylation simultaneously with kinetic parameters as

274 follows: k_{cat1}, 9.15 ± 0.08 s^{-1}, k_{cat2}, 37.8 ± 1.2 s^{-1}, K_{m1}, 0.127 ± 0.002 mM, and K_{m2}, 26.1 ± 0.7 mM
275 (**Figure 4c**). The kinetic parameter k_{cat1}/K_{m1}, which is comparable to k_{cat}/K_m of the Michaelis-Menten
276 equation, was 71.9 s^{-1} mM^{-1}. The transglucosylation ratio, $r_{tg} = v_{tg} / v_{ag}$, followed the equation 4 with
277 K_{TG} of 6.30 ± 0.08 mM, at which $v_h = v_{tg}$ is obtained (**Figure 4d**). In the lower pNP β-Glc
278 concentrations (≤ 0.6 mM), the reaction velocity for aglycone release followed Michaelis-Menten
279 equation with k_{cat}/K_m (72.2 s^{-1} mM^{-1}), which was consistent with k_{cat1}/K_{m1}. Compared with k_{cat1}/K_{m1} of
280 pNP β-Glc, k_{cat}/K_m of pNP β-D-Fuc, 81.3 s^{-1} mM^{-1}, was similar, but those of pNP β-Gal, pNP β-Man,
281 and pNP β-Xyl were only 1.9%, 1.1%, and 0.40% that of pNP β-Glc, respectively (**Table 2**).
282
283 **Specificity to various β-glucosides**
284 Reaction velocity was measured with 2 mM of various β-glucosides (**Table 2**). Reaction
285 velocities towards helicin, 4MU β-Glc, and scopolin were high, but those towards phenyl β-glucoside,
286 arbutin, salicylic acid β-glucoside, phlorizin, methyl β-glucoside, and octyl β-glucoside were low or
287 undetectable. k_{cat}/K_m for helicin, 160 s^{-1} mM^{-1}, was the highest of the tested substrates, followed by
288 those of scopolin (104 s^{-1} mM^{-1}) and 4MU β-Glc (97 s^{-1} mM^{-1}). Thus, the scopolin hydrolytic activity
289 of AtBGlu42, predicted in the plant reverse genetics studies (Zamioudis, Hanson and Pieterse 2014;
290 Stringlis *et al.* 2018), was enzymatically confirmed.
291
292 **Specificity to β-glucooligosaccharides**
293 AtBGlu42 hydrolyzed various β-glucobioses (**Table 2**). Among the disaccharides,
294 laminaribiose (Glcβ1-3Glc) was the best substrate in terms of k_{cat}/K_m, 94.6 s^{-1} mM^{-1}, followed by
295 sophorose (Glcβ1-2Glc, 13.8 s^{-1} mM^{-1}). Those of cellobiose (Glcβ1-4Glc), gentiobiose (Glcβ1-6Glc),
296 and neotrehalose (Glcβ1-α1Glc) were very low or undetectable. The highest k_{cat}/K_m was for
297 laminaribiose of laminarioligosaccharides, but for cellotriose (85.5 s^{-1} mM^{-1}) of cellooligosaccharides.
298 The decrease in k_{cat}/K_m values with increasing degree of polymerization above that of the best
299 substrates was milder with cellooligosaccharides than with laminarioligosaccharides.
300

301 **Structural analysis**

302 For the crystal structure analysis, the N-terminal tag of the recombinant AtBGlu42 was
303 removed by cleavage with thrombin and column chromatographies. The untagged protein possesses
304 two extra residues, Gly-Ser, before Met1 of the registered sequence. The specific activity of the
305 untagged enzyme (56.2 kDa) was 16.9 ± 0.3 U mg^{-1}, corresponding to 15.8 ± 0.3 s^{-1}, which is close to
306 that of the recombinant enzyme without removal of the N-terminal region. The structure of the
307 untagged AtBGlu42 was determined at 1.7 Å resolution by X-ray crystallography (**Table 1**). Because
308 of poor electron density, the N-terminal Gly, nine C-terminal residues Asp482–Glu490, and a part of
309 β→α loop 6 Lys328–Glu331 were not built. The overall structure of AtBGlu42 was a $(\beta/\alpha)_8$-barrel, as
310 are all known GH1 enzyme structures (**Figure 5a**). An extra N-terminal α-helix was located suitably
311 to interact with the small domain composed of β→α loops 5 and 6, in which a three-stranded β-sheet
312 consisting of β5′, β6′, and β6″ was observed. This extra α-helix was not found in the other GH1
313 structures, and the other *A. thaliana* and rice GH1 homologues, including Os1BGlu4, do not have the
314 amino acid sequence corresponding to the extra α-helix. The putative catalytic residues of AtBGlu42
315 (Glu183 and Glu388 on β-strands 4 and 7, respectively) and the residues interacting with the glucose
316 moiety in subsite −1 (Gln35, His137, Asn182, Tyr317, Glu388, Trp437, Glu444, Trp445, and Phe453)
317 are positioned similarly to those of known GH1 β-glucosidase structures (**Figure 6a**). Electron density
318 for four molecules of glycerol used as a cryoprotectant was observed. Two glycerol molecules were
319 located on the β→α loop 1 (**Figure 5a**). The other two were located in subsites −1 and +1 with
320 possible hydrogen bonds with Gln35, His137, Asn182, Glu183, Glu388, Glu444, and Trp445 (**Figure**
321 **5b**). The glycerol in subsite −1 was placed in the position corresponding to 2-C to 4-C of the substrate
322 analogous inhibitors (glucoimidazole and 1-deoxynojirimycin) bound to subsite −1 in other GH1 β-
323 glucosidases: Os3BGlu7 (PDB entry, 7BZM) (Nutho *et al.* 2020) and Tmari_1862 (PDB entry, 2CES)
324 (Gloster *et al.* 2006) (**Figure 6a**). Three O atoms of the glycerol were at the same positions of 2-O and
325 3-O of the inhibitors, but not at that of 4-O.

326 We compared the structures of AtBGlu42 with the Os3BGlu7 (Chuenchor *et al.* 2011) to
327 explore the structure involved in binding of laminaribiose and cellooligosaccharides. The acid/base

328 catalyst Glu183 of AtBGlu42 is at the same position as Glu176 (Gln in the structure) of Os3BGlu7-
329 laminaribiose complex to form hydrogen bonds with 2-O and 3-O (glucosidic oxygen) of the glucose
330 moiety at subsite +1, but Glu246 equivalent to Asn245 of Os3BGlu7 is far from 1-O. Instead, Asp244
331 is suitably located to form a hydrogen bond with 1-O, while corresponding Asp243 of Os3BGlu7 is
332 differently oriented toward Arg178 **(Figure 6b and c**). In the Os3BGlu7-cellopentaose complex, three
333 residues, Asn245 on β→α loop 5, Tyr341 and Trp358 on β→α loop 6, are placed to interact with
334 cellopentaose in subsites +1 to +4 **(Figure 6d**). The corresponding residues in AtBGlu42, Glu246,
335 Arg342, and Trp360, are placed in the equivalent positions. The Glu246 side chain, differently
336 oriented from Asn245 of Os3BGlu7, is also in a good position to form subsite +2 and no obvious
337 steric hindrance at subsite +1 and +2 is observed, but the side chain of Arg342 seems to occupy the
338 space for subsite +3 to cause a steric hindrance upon binding to cellooligosaccharide in subsite +3
339 **(Figure 6e**). The distance between Arg342 and the glucosyl residue at subsite +3 is too short (1.7 Å
340 between Arg342 Nη1 and Glc 5-C; 2.5 Å between Arg342 Nη2 and Glc 6-C).
341
342 **Analysis of Arg342 mutants**
343 The role of Arg342 in cellooligosaccharide selection was investigated through analysis of the
344 activity of Arg342 mutants, R342A and R342Y, on cellooligosaccharide (**Table 3**). The k_{cat}/K_m values
345 of R342A and R342Y for cellobiose were 72% and 5.4% of that of the wild type, respectively. Those
346 for pNP β-Glc were 52% and 6.9% of the wild type, respectively. Greater reductions in R342Y than in
347 R342A were observed for all the tested cellooligosaccharides, mainly because of this decrease in k_{cat}.
348 The values of k_{cat} in R342Y for pNP β-Glc and cellooligosaccharides decreased to 12-14% of those of
349 the wild type with unknown reason. The cellotriose preference of the wild type was not observed in
350 the mutants. The highest k_{cat}/K_m of R342A and R342Y was to cellopentaose and cellotetraose,
351 respectively. The ratio of k_{cat}/K_m of cellotetraose over cellotriose was 0.52 in the wild type, but 1.2 and
352 2.7 in R342A and R342Y, respectively. Subsite affinities were calculated (**Table 4**). Affinity of subsite
353 +3 was negative in the wild type (−1.64 kJ mol^{-1}), but positive affinity (0.470 kJ mol^{-1} and 2.54 kJ

354 mol^{-1}, respectively) was obtained in R342A and R342Y. The negative affinity observed at subsite +4
355 in the wild type was also changed to favorable or less unfavorable by the Arg342 mutations.
356
357 **Docking simulation of AtBGlu42 and scopolin**
358 Binding mode of scopolin in AtBGlu42 was predicted by docking simulation using AutoDock
359 Vina. The binding energy of water-free AtBGlu42 and scopolin was calculated to be −37.2 kJ mol^{-1}.
360 In the obtained structure, the glucosyl residue of scopolin was accommodated in subsite −1, as
361 observed in other GH1 complex structures, with possible interactions to Gln35, His137, Asn182,
362 Tyr317, Glu388, Trp437, Glu444, Trp445, and Phe453 (**Figure 7**). The umbelliferyl plane of the
363 aglycone part was placed to face Trp360 for the possible stacking interaction. The 6-methoxy group of
364 scopolin was at a distance from Phe197 (3.8 Å between C–C, 3.4 Å between C–O).
365
366 **Discussion**
367 Plants possess numerous GH1 β-glucosidase isoenzymes, and they are involved in various
368 biological processes including development and biotic/abiotic stress responses. The functional
369 characterization of the enzymes is important to address their physiological functions and
370 understanding the molecular basis of the enzymatic functions makes it possible to predict enzyme
371 function more accurately. AtBGlu42, a putative cytosolic GH1 β-glucosidase, was previously
372 implicated, by the reverse genetic approach, to be responsible for the hydrolysis of scopolin in roots to
373 release an iron-chelating agent scopoletin into the rhizosphere under iron-deficient conditions
374 (Zamioudis, Hanson and Pieterse 2014; Stringlis *et al.* 2018). This suggests that AtBGlu42 is highly
375 specific to this β-glucoside, but the homologous enzyme Os1BGlu4 from rice plant efficiently
376 catalyzes the hydrolysis of cello- and laminari-oligosaccharides in addition to the β-glucosides of
377 plant secondary metabolites (Rouyi *et al.* 2014). In this study, the activity of AtBGlu42 on scopolin
378 and various substrates, including these oligosaccharides, was clarified along with the associated
379 protein structure using the enzyme produced by recombinant expression.

380 Recombinant AtBGlu42 was the most active at a neutral pH, similar to Os1BGlu4 (Rouyi *et al.*
381 2014). This optimum pH is suitable for the action in the cytoplasm since its pH in plants is neutral
382 (Shen *et al.* 2013). The homologous rice protein Os1BGlu4 possesses no signal peptide in its
383 precursor sequence, and its cytoplasmic localization has been demonstrated in maize protoplasts by
384 transient expression of the Os1BGlu4-GFP fusion protein (Rouyi *et al.* 2014). AtBGlu42 is also
385 presumably located in cytoplasm. In addition to the neutral pH optima, the substrate specificity of
386 AtBGlu42 on those secondary metabolite glucosides and the short-chain oligosaccharides may be
387 reasonable as a cytosolic enzyme.
388 The hydrolytic activity of AtBGlu42 on scopolin was confirmed in this study. AtBGlu42
389 hydrolyzed scopolin with the second highest k_{cat}/K_m of the tested substrates (**Table 2**). The docking
390 simulation of AtBGlu42 and scopolin predicted van der Waals interaction of Phe197 with the 6-
391 methoxy group and the stacking interaction of Trp360 with the umbelliferyl moiety, respectively
392 (**Figure 7**). AtBGlu42 acted on 4MU β-Glc with almost the same k_{cat} and K_m as on scopolin,
393 suggesting that the interaction of Phe197 with the 6-methoxy group of scopolin is comparable to the
394 interaction between Trp360 and the 4-methyl group, which is expected to be formed assuming that the
395 umbelliferyl ring of 4MU β-Glc binds in the same way as predicted for scopolin. The k_{cat}/K_m value of
396 AtBGlu42 was 61-fold or much higher than those of other Arabidopsis GH1 β-glucosidase
397 isoenzymes, AtBGlu21, AtBGlu22, and AtBGlu23, which are enzymes known to hydrolyze scopolin
398 (Ahn *et al.* 2010). These isoenzymes have Lys and Ala residues at the position corresponding to
399 Phe197 and Trp360 in AtBGlu42, respectively. Thus, Phe197 and Trp360 are presumably key residues
400 for high activity to scopolin. Of the 40 Arabidopsis GH1 isoenzymes, three (AtBGlu44, AtBGlu45,
401 and AtBGlu46) have a set of Phe and Trp residues in addition to AtBGlu42, and 12 of the 40
402 isoenzymes have Trp residue, which is thought to be important for stacking with the δ-lactone ring.
403 Since many GH1 enzymes are highly active for 4MU β-Glc, and Trp residues are conserved in more
404 than 80% of the approximately 300 characterized GH1 enzymes (mainly from bacteria), many GH1
405 enzymes may also act on scopolin efficiently.

406 Among the tested β-glucosides, AtBGlu42 showed the highest k_{cat}/K_m to helicin (2-*O*-
407 glucosylbenzaldehyde), higher than scopolin, and considerable activity to pNP β-Glc. On the other
408 hand, activity on phenyl β-glucoside, arbutin (*p*-hydroxyphenyl β-glucoside), salicylic acid β-
409 glucoside (2-*O*-glucosylbenzoic acid) and phlorizin were very low or undetectable (**Table 2**). The
410 aldehyde group of helicin could be equivalent to 6-methoxy group of scopolin, the carbon of which
411 was in a distance for hydrophobic interaction with Phe197 in the modelled complex structure (**Figure**
412 **7**). GH1 β-glucosidase from *Agrobacterium tumefaciens*, SghA, acts on salicylic acid β-glucoside
413 through the formation of hydrogen bonds between the carboxy group and His193, corresponding to
414 Phe197 of AtBGlu42 (Wang *et al.* 2019). Charged groups or groups bulkier than the aldehyde group
415 on this position of the phenyl group may not be accepted because of unfavorable interactions with
416 Phe197.
417 AtBGlu42 showed activity not only on β-glucosides but also on other β-glycosides (**Table 2**).
418 Among the tested pNP β-glycosides, pNP β-Glc and pNP β-D-Fuc (6-deoxy-D-galactoside) were good
419 substrates as observed in several plant β-glucosidases (Ketudat Cairns and Esen 2010). The much
420 lower k_{cat}/K_m values to pNP β-Gal and pNP β-Xyl were mainly due to high K_m values. pNP β-Man was
421 also a poor substrate but mainly due to low k_{cat}. This indicates that pNP β-Man is not favorable for the
422 formation of the transition state from the Michaelis complex. The axial 2-OH group of β-mannoside is
423 probably not suitable for the stabilization of the transition state in the active site of AtBGlu42.
424 Efficient activity on β-mannosides requires the ability to bind their transition state in the $B_{2,5}$ boat
425 conformation favored for mannoside hydrolysis rather than the $^4H_3$ half-chair or ^4E envelope-like
426 transition state of glucosides (Tankrathok *et al.* 2015). Subtle differences in the active site amino acid
427 residue positions apparently enable this in some GH1 enzymes, but not others, including AtBGlu42.
428 In the reaction with pNP β-Glc, AtBGlu42 catalyzed both hydrolysis and transglucosylation
429 with K_{TG} of 6.3 mM. The reaction rates for hydrolysis, transglucosylation, and aglycone release
430 obeyed well the reaction equations obtained based on the reaction scheme for retaining glycosidase
431 well. The transglucosylation products including pNP β-cellobioside were observed, as observed for
432 Os1BGlu4 (Rouyi *et al.* 2014), Os3BGlu7 (Opassiri *et al*. 2004), and HvBII (Hrmova *et al*. 1998).

433 The transglycosylation was not observed with the other pNP β-glycosides under the reaction
434 conditions, suggesting that only β-glucoside can be bound in the suitable position for
435 transglucosylation even though it must take a few binding modes to produce the transglycosylation
436 products.

437 AtBGlu42 also hydrolyzed various β-glucooligosaccharides. In terms of k_{cat}/K_m, laminaribiose
438 was the best substrate among the tested disaccharides and also among laminarioligosaccharides, while
439 cellotriose was the best of cellooligosaccharides. The disaccharide preference among
440 laminarioligosaccharides is often observed in plant β-glucosidases (Opassiri *et al.* 2004; Seshadri *et*
441 *al.* 2009; Opassiri *et al.* 2010; Rouyi *et al.* 2014). The crystal structure of the Os3BGlu7-laminaribiose
442 complex indicates that the reducing-terminal β-glucosyl residue of laminaribiose binds to subsite +1 in
443 a similar position to the corresponding glucosyl unit of cellooligosaccharide, but the faces of their
444 glucosyl units toward Trp358 were opposite to each other. Therefore 1-O of the laminaribiose is
445 placed on the bottom of the cleft, close to 6-O of the Glc moiety in the cellooligosaccharide, with a
446 possible hydrogen bond to Asn245 (Chuenchor *et al.* 2011) (**Figure 6b and d**). AtBGlu42 seems to
447 possess the residues to accommodate laminaribiose at the similar position in the protein structure
448 (**Figure 6c**). In the binding of longer laminarioligosaccharides, the corresponding Glc moiety needs to
449 take another unfavorable binding mode in subsite +1 to avoid the steric hindrance. This could explain
450 the disaccharide preference of laminarioligosaccharides **(Table 2)**.

451 The cellotriose preference of AtBGlu42 is caused by the negative affinity in subsite +3, mainly
452 due to the steric hindrance with the Arg342 side chain. Structural comparison with the Os3BGlu7-
453 cellopentaose complex clearly shows that the Arg342 side chain occupies the substrate binding cleft to
454 cause steric hindrance in subsite +3 upon binding to cellooligosaccharides, while the corresponding
455 Tyr341 in Os3BGlu7 has stacking interaction onto the glucosyl residue in subsite +3 (**Figure 6d and**
456 **e**). The role of the Arg342 was supported by the results of its mutants, R342A and R342Y, both of
457 which showed clear increases in the affinity in subsite +3 (**Table 4**). Particularly the affinity of R342Y
458 (2.5 kJ mol^{-1}) is similar to that of Os3BGlu7 (2.2 kJ mol^{-1}) (Opassiri *et al.* 2004), implying that the
459 Tyr substitution sets the binding site as in Os3BGlu7.

460 Arg342 is responsible for the cellotriose preference in AtBGlu42, but careful consideration is
461 required to predict the cellooligosaccharide preference of other GH1 β-glucosidases, because some
462 enzymes possessing Arg residues in the position do not exhibit a cellotriose preference. The
463 homologous rice enzyme, Os1BGlu4, possesses equivalent Arg334, but exhibits the highest preference
464 for cellotetraose of cellooligosaccharides with 1.9-fold higher k_{cat}/K_m than cellotriose (Rouyi *et al.*
465 2014). Its three-dimensional structure is not known yet, but the lack of the trisaccharide preference is
466 possibly caused by a different orientation of the Arg334 side chain. One possible reason for it is
467 interaction of Arg334 with its neighboring Glu336, corresponding to Val344 in AtBGlu42. Another
468 example is a metagenomic β-glucosidase Td2F2. This enzyme, with 41% sequence identity with
469 AtBGlu42, possesses the corresponding Arg313 (Matsuzawa *et al.* 2016), but it shows 1.5-fold higher
470 k_{cat}/K_m to cellotetraose than to cellotriose (Uchiyama, Miyazaki and Yaoi 2013). The structural
471 comparison revealed that the Arg residues in the both structures are placed in the same position in the
472 local small domain structure composed of β→α loops 5 and 6, which contain β5′–α5′ and β6′–β6″,
473 respectively. The three strands β5′, β6′, and β6″ comprise β-sheet in the small domain as in other GH1
474 enzymes, and the Arg residues are similarly located on β6″ in a three-stranded β-sheet (**Figure 8a and**
475 **b**) (Matsuzawa *et al.* 2016). However, the Arg side chains in AtBGlu42 and Td2F2 are placed
476 differently. One reason is different interactions with surrounding residues (**Figure 8a and b**). Arg342
477 of AtBGlu42 is in the position to form a salt bridge with Glu246 on β→α loop 5 (**Figure 8a**). In the
478 Td2F2 structure, the corresponding residue is Thr225, and Arg313 has another interaction with the
479 carbonyl oxygen of Arg297 in the N-terminal stretch of β6′, to place the side chain of Arg313 out of
480 the glucosyl binding site (**Figure 8b**). The other reason is a shift of the small local domain location in
481 the $(\beta/\alpha)_8$-barrel structure. The small domains are placed a bit differently while the β5, α5, β6, and α6
482 composing the $(\beta/\alpha)_8$-barrel are strictly placed in the same positions (**Figure 8c**). The small domain of
483 AtBGlu42 is located closer to α6, compared with that of Td2F2. The distance between Cα atoms of
484 Pro255 and Pro234, located at the N-terminus of α5′ of AtBGlu42 and Td2F2, respectively, is 7.0 Å in
485 the superimposed structures. That of the Arg residues is 2.9 Å. The different placement of the local
486 domains is stabilized by the interaction with the surrounding structures (**Figure 8d and e**). In the

Td2F2 structure, two salt bridges connecting α5′ to α6, and α5 to an α-helix adjacent to α5′ are observed (Asp244–Arg338 and Arg249–Asp263, respectively). In the structure of AtBGlu42, those salt bridges are not found, but a hydrophobic core, consisting of Trp247, Ile265, Lys370, and Tyr374, is formed between the small domain and α6. Mainly because of these differences, the small domains in the shifted positions each other are stabilized, providing the different surrounding circumstances to the Arg side chains.

In this study, we have elucidated the biochemical functions and structure of AtBGlu42, responsible to hydrolyze scopolin in Arabidopsis. Recombinant AtBGlu42 showed high activities not only toward scopolin, but also towards other β-glucosides such as helicin, laminaribiose, and cellotriose. The preference for chain-length of cellooligosaccharides was understood by possible steric hindrance caused by Arg342 on the β→α loop 6 at subsite +3. Combining the mutational study of AtBGlu42, we postulate β→α loop 6, including amino acid residue corresponding to Arg342 of AtBGlu42, is the key determinant of specificity to cellooligosaccharide chain-length in GH1 enzymes. Furthermore, structural comparison of GH1 β-glucosidases suggests that a mechanism for diversification of substrate binding site through variation in structures that do not directly interact with the substrates.

Acknowledgements

We thank Mr. Yusuke Takada of the DNA sequencing facility of the Research Faculty of Agriculture, Hokkaido University for assistance with DNA sequence analysis; and Ms. Nozomi Takeda of the Global Facility Center, Hokkaido University for the amino acid analysis. X-ray diffraction experiments were conducted at SPring-8 (proposals 2021A2756). We thank the beamline staff of SPring-8 for their assistance.

Data availability

The data underlying this article are available in the article.

514 **Author contributions**

515 S.H. conceived and designed the experiments, performed the biochemical experiments, determined
516 the protein structure, and wrote the paper. W.S. conceived and designed the experiments, determined
517 the protein structure, and wrote the paper. J.Y. determined the protein structure and wrote the paper.
518 H.Matsuura prepared the substrate and wrote the paper. J.R.K.C. wrote the paper. M.Y. determined the
519 protein structure and wrote the paper. H.Mori conceived and designed the experiments and wrote the
520 paper.

521

522 **Funding**

523 This research was supported by Platform Project for Supporting Drug Discovery and Life Science
524 Research (Basis for Supporting Innovative Drug Discovery and Life Science Research; BINDS) from
525 AMED under grant no. JP18am00101071 and JP19am0101093.

526

527 **Disclosure statement**

528 No potential conflict of interest was reported by the authors.

529

530 **References**

531 Adams PD, Afonine PV, Bunkóczi G *et al*. *PHENIX*: a comprehensive Python-based system for
532 macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:213-221.

533 Ahn YO, Shimizu B, Sakata K *et al*. Scopolin-hydrolyzing β-glucosidases in roots of Arabidopsis.
534 *Plant Cell Physiol* 2010;**51**:132-143.

535 Baba SA, Vishwakarma RA, Ashraf N. Functional characterization of *Cs*BGlu12, a β-glucosidase
536 from *Crocus sativus*, provides insights into its role in abiotic stress through accumulation of
537 antioxidant flavonols. *J Biol Chem* 2017;**292**:4700-4713.

538 Babcock GD, Esen A. Substrate specificity of maize β-glucosidase. *Plant Sci* 1994;**101**:31-39.

539 Barleben L, Panjikar S, Ruppert M *et al*. Molecular architecture of strictosidine glucosidase: the
540 gateway to the biosynthesis of the monoterpenoid indole alkaloid family. *Plant Cell*
541 2007;**19**:2886-2897.

542 Chen VB, Arendall WB III, Headd JJ *et al*. *MolProbity*: all-atom structure validation for
543 macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:12-21.
544 Chuenchor W, Pengthaisong S, Robinson RC *et al*. The structural basis of oligosaccharide binding by
545 rice BGlu1 beta-glucosidase. *J Struct Biol* 2011;**173**:169-179.
546 Emsley P, Cowtan K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr D Biol*
547 *Crystallogr* 2004;**60**:2126-2132.
548 Gloster TM, Roberts S, Perugino G *et al*. Structural, kinetic, and thermodynamic analysis of
549 glucoimidazole-derived glycosidase inhibitors. *Biochemistry* 2006;**45**:11879-11884.
550 Grynkiewicz G, Achmatowicz O, Hennig J *et al.* Synthesis and characterization of the salicylic acid β-
551 D-glucopyranoside. *Pol J Chem* 1993;**67**:1251-1254.
552 Henrissat B, Callebaut I, Fabrega S *et al*. Conserved catalytic machinery and the prediction of a
553 common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci USA*
554 1995;**92**:7090-7094.
555 Himeno N, Saburi W, Wakuta S *et al*. Identification of rice β-glucosidase with high hydrolytic activity
556 towards salicylic acid β-D-glucoside. *Biosci Biotechnol Biochem* 2013;**77**:934-939.
557 Hiromi K, Nitta Y, Numata C *et al.* Subsite affinities of glucoamylase: examination of the validity of
558 the subsite theory. *Biochim Biophys Acta* 1973;**302**:362-375.
559 Hösel W, Tober I, Eklund SH *et al*. Characterization of β-glucosidases with high specificity for the
560 cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench seedlings. *Arch Biochem*
561 *Biophys* 1987;**252**:152-162.
562 Hrmova M, MacGregor EA, Biely P *et al*. Substrate binding and catalytic mechanism of a barley β-D-
563 Glucosidase/(1,4)-β-D-glucan exohydrolase. *J Biol Chem* 1998;**273**:11134-11143.
564 Hua Y, Sansenya S, Saetang C *et al*. Enzymatic and structural characterization of hydrolysis of
565 gibberellin A4 glucosyl ester by a rice β-D-glucosidase. *Arch Biochem Biophys* 2013;**537**:39-
566 48.
567 Jenkins J, Lo Leggio L, Harris G *et al*. β-Glucosidase, β-galactosidase, family A cellulases, family F
568 xylanases and two barley glycanases form a superfamily of enzymes with 8-fold β/α

569 architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands
570 four and seven. *FEBS Lett* 1995;**362**:281-285.
571 Kabsch W. *XDS*. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:125-132.
572 Ketudat Cairns JR, Esen A. β-Glucosidases. *Cell Mol Life Sci* 2010;**67**:3389-3405.
573 Ketudat Cairns JR, Pengthaisong S, Luang S *et al.* Protein-carbohydrate interactions leading to
574 hydrolysis and transglycosylation in plant glycoside hydrolase family 1 enzymes. *J Appl*
575 *Glycosci* 2012;**59**:51-62.
576 Ketudat Cairns JR, Mahong B, Baiya S *et al*. β-Glucosidases: Multitasking, moonlighting or simply
577 misunderstood? *Plant Sci* 2015;**241**:246-259.
578 Kobayashi M, Hondoh H, Mori H *et al*. Calcium ion-dependent increase in thermostability of dextran
579 glucosidase from *Streptococcus mutans*. *Biosci Biotechnol Biochem* 2011;**75**:1557-1563.
580 Lee KH, Piao HL, Kim HY *et al*. Activation of glucosidase via stress-induced polymerization rapidly
581 increases active pools of abscisic acid. *Cell* 2006;**126**:1109-1120.
582 Lombard V, Golaconda Ramulu H, Drula E *et al.* The carbohydrate-active enzymes database (CAZy)
583 in 2013. *Nucleic Acids Res* 2014;**42**:D490-495.
584 McCoy AJ, Grosse-Kunstleve RW, Adams PD *et al*. *Phaser* crystallographic software. *J Appl*
585 *Crystallogr* 2007;**40**:658-674.
586 Matsuzawa T, Jo T, Uchiyama T *et al*. Crystal structure and identification of a key amino acid for
587 glucose tolerance, substrate specificity, and transglycosylation activity of metagenomic β-
588 glucosidase Td2F2. *FEBS J* 2016;**283**:2340-2353.
589 Matthews BW. Solvent content of protein crystals. *J Mol Biol* 1968;**33**:491-497.
590 Nutho B, Pengthaisong S, Tankrathok A *et al*. Structural basis of specific glucoimidazole and
591 mannoimidazole binding by Os3BGlu7. *Biomolecules* 2020;**10**:907.
592 Opassiri R, Hua Y, Wara-Aswapati O *et al*. β-Glucosidase, exo-β-glucanase and pyridoxine
593 transglucosylase activities of rice BGlu1. *Biochem J* 2004;**379**:125-131.
594 Opassiri R, Pomthong B, Onkoksoong T *et al*. Analysis of rice glycosyl hydrolase family 1 and
595 expression of Os4bglu12 β-glucosidase. *BMC Plant Biol* 2006;**6**:33.

596 Opassiri R, Maneesan J, Akiyama T *et al*. Rice Os4BGlu12 is a wound-induced β-glucosidase that
597 hydrolyzes cell wall-β-glucan-derived oligosaccharides and glycosides. *Plant Sci*
598 2010;**179**:273-280.
599 Roepke J, Bozzo GG. *Arabidopsis thaliana* β-glucosidase BGLU15 attacks flavonol 3-*O*-β-glucoside-
600 7-*O*-α-rhamnosides. *Phytochemistry* 2015;**109**:14-24.
601 Rouyi C, Baiya S, Lee SK *et al*. Recombinant expression and characterization of the cytoplasmic rice
602 β-glucosidase Os1BGlu4. *PLoS One* 2014;**9**:e96712.
603 Rye CS, Withers SG. Glycosidase mechanisms. *Curr Opin Chem Biol* 2000;**4**:573-580.
604 Saburi W, Mori H, Saito S *et al.* Structural elements in dextran glucosidase responsible for high
605 specificity to long chain substrate. *Biochim Biophys Acta* 2006;**1764**:688-698.
606 Saburi W, Kobayashi M, Mori H *et al.* Replacement of the catalytic nucleophile aspartyl residue of
607 dextran glucosidase by cysteine sulfinate enhances transglycosylation activity. *J Biol Chem*
608 2013;**288**:31670-31677.
609 Seshadri S, Akiyama T, Opassiri R *et al.* Structural and enzymatic characterization of Os3BGlu6, a
610 rice β-glucosidase hydrolyzing hydrophobic glycosides and (1→3)- and (1→2)-linked
611 disaccharides. *Plant Physiol* 2009;**151**:47-58.
612 Shen J, Zeng Y, Zhuang X *et al*. Organelle pH in the *Arabidopsis* endomembrane system. *Mol Plant*
613 2013;**6**:1419-1437.
614 Stringlis IA, Yu K, Feussner K *et al*. MYB72-dependent coumarin exudation shapes root microbiome
615 assembly to promote plant health. *Proc Natl Acad Sci USA* 2018;**115**:E5213-5222.
616 Sue M, Ishihara A, Iwamura H. Purification and characterization of a hydroxamic acid glucoside β-
617 glucosidase from wheat (*Triticum aestivum* L.) seedlings. *Planta* 2000;**210**:432-438.
618 Tankrathok A, Iglesias-Fernández J, Williams RJ *et al*. A single glycosidase harnesses different
619 pyranoside ring transition state conformations for hydrolysis of mannosides and glucosides.
620 *ACS Catal* 2015;**5**:6041-6051.
621 Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring
622 function, efficient optimization, and multithreading. *J Comput Chem* 2010;**31**:455-461.

623 Uchiyama T, Miyazaki K, Yaoi K. Characterization of a novel β-glucosidase from a compost
624 microbial metagenome with strong transglycosylation activity. *J Biol Chem* 2013;**288**:18325-
625 18334.
626 Wakuta S, Hamada S, Ito H *et al*. Identification of a β-glucosidase hydrolyzing tuberonic acid
627 glucoside in rice (*Oryza sativa* L.). *Phytochemistry* 2010;**71**:1280-1288.
628 Wakuta S, Hamada S, Ito H *et al.* Comparison of enzymatic properties and gene expression profiles of
629 two tuberonic acid glucoside β-glucosidases from *Oryza sativa* L. *J Appl Glycosci*
630 2011;**58**:67-70
631 Wang C, Ye F, Chang C *et al.* Agrobacteria reprogram virulence gene expression by controlled
632 release of host-conjugated signals. *Proc Natl Acad Sci USA* 2019;**116**:22331-22340.
633 Xia L, Ruppert M, Wang M *et al*. Structures of alkaloid biosynthetic glucosidases decode substrate
634 specificity. *ACS Chem Biol* 2012;**7**:226-234.
635 Xu Z, Escamilla-Treviño LL, Zeng L *et al*. Functional genomic analysis of *Arabidopsis thaliana*
636 glycoside hydrolase family 1. *Plant Mol Biol* 2004;**55**:343-367.
637 Zamioudis C, Hanson J, Pieterse CMJ. β-Glucosidase BGLU42 is a MYB72-dependent key regulator
638 of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in
639 *Arabidopsis* roots. *New Phytol* 2014;**204**:368-379.
640

641 **Figure captions**

642 **Graphical abstract**

643 The cellotriose specificity of AtBGlu42 was attributed to R342, and the chain length specificity was
644 modified by mutation of R342.

645

646 **Figure 1.** Reaction scheme for the reaction with pNP β-Glc catalyzed by AtBGlu42.
647 E and E-Glc represent AtBGlu42 (enzyme) and the glucosyl-enzyme intermediate, respectively. pNP-
648 Glc_2 is the transglucosylation product.

649

650 **Figure 2.** SDS-PAGE analysis of purified recombinant AtBGlu42.

651 Lane 1, protein size marker; Lane 2, purified tagged AtBGlu42 (1 μg); Lane 3, purified untagged
652 AtBGlu42 (1 μg). Protein was stained with CBB. Molecular masses of the standard proteins are
653 shown on the left. Recombinant tagged AtBGlu42 (70 kDa) and untagged AtBGlu42 (56 kDa) are
654 indicated by an arrow.

655

656 **Figure 3.** Effects of pH and temperature on activity and stability of recombinant AtBGlu42.

657 (a) Relative activity at various pH (pH 4.1-11.0). (b) Relative activity at various temperatures (25-
658 60 °C). (c) Residual activity after AtBGlu42 was kept in the pH range at 4 °C for 24 h. (d) Residual
659 activity after AtBGlu42 was kept at the temperatures at pH 7.0 for 15 min. Values and error bars are
660 average and standard deviation of three independent experiments, respectively.

661

662 **Figure 4.** Reaction of recombinant AtBGlu42 with pNP β-Glc.

663 (a, b) TLC analysis of the reaction of AtBGlu42 (2.27 μM) and 8 mM pNP β-Glc in 50 mM sodium
664 phosphate buffer (pH 6.8) at 30 °C. Reaction products with pNP group were detected under UV light
665 (a), and those containing sugar were detected by staining (b). (c) The s-*v* plots of the reaction of with
666 pNP β-Glc. Velocity for aglycone release (circle), hydrolysis (triangle), and transglucosylation
667 (square) are shown. (d) Transglucosylation ratio ($r_{tg} = v_{tg} / v_{ag}$) against pNP β-Glc concentration.
668 Values and error bars are average and standard deviation of three independent experiments,
669 respectively. Theoretical lines according to the kinetic scheme are drawn.

670

671 **Figure 5.** Crystal structure of AtBGlu42.

672 (a) Ribbon diagram of overall AtBGlu42. The structure is shown in rainbow coloring from blue (N-
673 terminal) to red (C-terminal). Four glycerol molecules are shown in ball and stick representation. The
674 extra N-terminal α-helix is located close to the small local domain composed β→α loops 5
675 (containing β-strand β5′ and α-helix α5′) and 6 (containing two β-strands β6′ and β6″). (b) Close up
676 view of the active site of AtBGlu42. Two glycerol molecules observed are represented with balls and

677 sticks. Water molecules are shown as red balls. Predicted hydrogen bonds are indicated by dotted
678 lines. The F_o−F_c omit electron density map for the glycerol molecules is shown as a mesh contoured
679 at 3.5 σ.
680
681 **Figure 6.** Comparison of structure of AtBGlu42 with GH1 β-glucosidases.
682 (a) Superimposition of the residues at subsite −1 of AtBGlu42 (pink) with Os3BGlu7 (pale green),
683 and Tmari_1862 (light purple). Inhibitors (glucoimidazole in Os3BGlu7 and Tmari_1862) are shown
684 in ball and stick representation. (b) The Os3BGlu7-laminaribiose complex (PDB entry, 3AHT)
685 (Chuenchor *et al.* 2011). (c) The corresponding part of AtBGlu42. Laminaribiose of the Os3BGlu7
686 complex superimposed with AtBGlu42 is shown. (d) The Os3BGlu7-cellopentaose complex (PDB
687 entry, 3F5K) (Chuenchor *et al.* 2011). (e) The corresponding part of AtBGlu42. Cellopentaose of the
688 Os3BGlu7 complex superimposed with AtBGlu42 is shown.
689
690 **Figure 7.** Docking simulation of scopolin onto AtBGlu42.
691 Scopolin is shown with balls and sticks. The amino acid residues surrounding it are shown behind
692 transparent surface in grey for carbon. The umbelliferyl plane faces Trp360 and 6-methoxy group is at
693 a distance of 3.4-3.8 Å from Phe197.
694
695 **Figure 8.** Comparison of the three-dimensional structures of AtBGlu42 and Td2F2.
696 (a) Arg342 and its surrounding structure of AtBGlu42. The ligand is cellopentaose from the
697 Os3BGlu7 complex. (b) Corresponding Arg313 and structure of Td2F2 (PDB entry, 3WH5)
698 (Matsuzawa *et al.* 2016). (c) Superposition of the small domain structures of AtBGlu42 (pink) and
699 Td2F2 (purple). The small local domain composed of β→α loops 5 and 6 containing β-sheet (β5′, β6′,
700 and β6″) is shown along with β-strands 5 and 6 and α-helices 5 and 6 of the $(\beta/\alpha)_8$-barrel. (d) View of
701 the same local structure of AtBGlu42 from the right side. The hydrophobic core is observed between
702 the small domain and α6. (e) The equivalent view of Td2F2. The hydrophobic core is not found, but
703 possible salt-bridges between Asp244–Arg338 and Arg249–Asp264 are.

Table 1. Summary of crystallization conditions, data collection, and refinement statistics.

	AtBGlu42-Apo
Data collection	
PDB entry	7F3A
Beamline	SPring-8 BL41XU
Space group	$P2_1$
Unit cell parameters a, b, c, (Å)	44.7, 93.1, 60.9
Unit cell parameters α, β, γ, (º)	90, 102.5, 90
Wavelength (Å)	1
Resolution range (Å)	50.0-1.70 (1.80-1.70)
R_{meas} (%)[a]	8.1 (84.0)
$CC_{1/2}$	(0.783)
<I/σ(I)>	15.1 (2.00)
Completeness (%)	99.9 (99.7)
Redundancy	6.76 (5.87)
Refinement	
No. reflection	53,471
R_{work}/R_{free} (%)[b]	16.0/18.0
No. of atoms	
Macromolecules	3,865
Ligand/ion	24
Water	402
B-factors (Å^2)	
Macromolecules	25
Ligand/ion	30.4
Water	34.6
RMSD from ideal	
Bond lengths (Å)	0.006
Bond angles (°)	0.796
Ramachandran	
Favored (%)	98.1
Allowed (%)	1.9
Outliers (%)	0

Values in parentheses are for the highest resolution shell.
a, $R_{meas} = \Sigma_{hkl} \{N(hkl) / [N(hkl) - 1]\}^{1/2} \Sigma_i | I_i(hkl) - \langle I(hkl)\rangle | / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $\langle I(hkl)\rangle$ and $N(hkl)$ are the mean intensity of a set of equivalent reflections and the multiplicity, respectively.
b, $R_{work} = \Sigma_{hkl} ||F_{obs}| - |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, R_{free} was calculated for 5% randomly selected test sets that were not used in the refinement.

Table 2. Reaction velocities and kinetic parameters of recombinant AtBGlu42 for various substrates.

Substrate	Reaction velocity [a] (s^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})
pNP β-D-glucopyranoside	12.3 ± 0.2	10.2 ± 0.0 [b]	0.142 ± 0.002 [b]	72.2 [b] (71.9 [c])
pNP β-D-fucopyranoside	12.0 ± 0.0	12.9 ± 0.4	0.158 ± 0.004	81.3
pNP β-D-galactopyranoside	2.39± 0.04	13.6 ± 0.4	9.93 ± 0.41	1.34
pNP β-D-mannopyranoside	0.203 ± 0.003	0.233 ± 0.014	0.290 ± 0.005	0.804
pNP β-D-xylopyranoside	0.258 ± 0.002	0.464 ± 0.008	1.62 ± 0.04	0.287
pNP β-D-cellobioside	8.99 ± 0.95	9.40 ± 0.20	0.0978 ± 0.0028	96.1
4-Methylumbelliferyl β-D-glucopyranoside	8.03 ± 0.09	8.03 ± 0.08	0.0827 ± 0.0022	97.1
Scopolin	8.78 ± 0.01	10.2 ± 0.02	0.0981 ± 0.0001	104
Helicin	10.2 ± 0.3	10.5 ± 0.03	0.0657 ± 0.0044	160
Salicylic acid β-D-glucopyranoside	0.316 ± 0.014	N.D.	N.D.	N.D.
Arbutin	< 0.0046	N.D.	N.D.	N.D.
Phlorizin	0.418 ± 0.011	N.D.	N.D.	N.D.
Phenyl β-D-glucopyranoside	0.0897 ± 0.0046	N.D.	N.D.	N.D.
Methyl β-D-glucopyranoside	< 0.0046	N.D.	N.D.	N.D.
Octyl β-D-glucopyranoside	0.0254 ± 0.0027	N.D.	N.D.	N.D.
Cellobiose	2.00 ± 0.01	6.42 ± 0.09	4.07 ± 0.04	1.58
Cellotriose	9.99 ± 0.62	10.8 ± 0.1	0.126 ± 0.002	85.5
Cellotetraose	9.34 ± 0.31	10.3 ± 0.3	0.231 ± 0.01	44.7
Cellopentaose	8.59 ± 0.20	10.2 ± 0.2	0.283 ± 0.02	35.9
Cellohexaose	8.33 ± 0.15	9.67 ± 0.5	0.314 ± 0.02	30.8
Laminaribiose	10.6 ± 0.09	11.1 ± 0.2	0.117 ± 0.004	94.6
Laminaritriose	8.07 ± 0.35	10.7 ± 1.0	0.636 ± 0.09	16.9
Laminaritetraose	2.98 ± 0.29	N.D.	N.D.	1.68
Sophorose	8.35 ± 0.10	11.9 ± 0.2	0.864 ± 0.022	13.8
Gentiobiose	0.0593 ± 0.0040	0.543 ± 0.073	16.9 ± 3.1	0.032
Neotrehalose	< 0.0046	N.D.	N.D.	N.D.

Data are average ± standard deviation for three independent experiments. N.D., not determined. a, Reaction velocity to 2 mM substrates. b, Kinetic parameter of Michaelis-Menten equation determined in the low substrate concentration range (≤ 0.6 mM). c, k_{cat1}/K_{m1} of the kinetic model of retaining glycosidases that catalyze both hydrolysis and transglycosylation.

Table 3. Kinetic parameters of recombinant AtBGlu42 mutants for cellooligosaccharides.

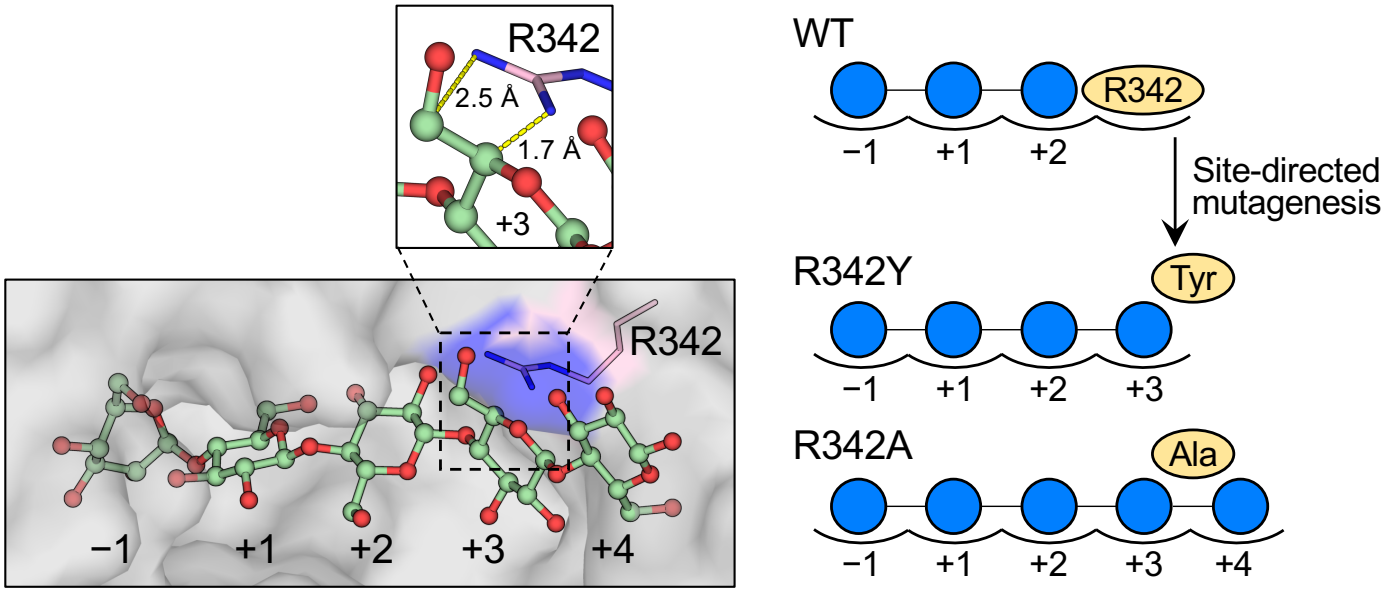
Substate	R342A			R342Y		
	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})
pNP β-D-glucopyranoside	7.89 ± 0.13 [a]	0.211 ± 0.008 [a]	37.5 [a]	1.35 ± 0.01 [a]	0.270 ± 0.005 [a]	4.98 [a]
Cellobiose	5.43 ± 0.16	4.82 ± 0.34	1.13	0.927 ± 0.044	10.8 ± 0.8	0.0855
Cellotriose	7.54 ± 0.13	0.136 ± 0.010	55.4	1.33 ± 0.03	0.323 ± 0.017	4.12
Cellotetraose	8.21 ± 0.05	0.122 ± 0.007	67.1	1.23 ± 0.01	0.109 ± 0.005	11.3
Cellopentaose	8.08 ± 0.11	0.107 ± 0.003	75.8	1.31 ± 0.04	0.127 ± 0.008	10.3
Cellohexaose	8.10 ± 0.22	0.151 ± 0.014	53.7	1.29 ± 0.03	0.147 ± 0.004	8.78

a, Kinetic parameter of Michaelis-Menten equation determined in the low substrate concentration range (≤ 0.6 mM).

Table 4. Subsite affinities for cellooligosaccharide.

	Subsite affinity (kJ mol^{-1})		
Subsite	Wild type	R342A	R342Y
−1	11.2 ± 0.1	11.6 ± 0.2	11.9 ± 0.2
+1	11.6 ± 0.1	11.0 ± 0.3	8.75 ± 0.27
+2	10.1 ± 0.0	9.83 ± 0.12	9.78 ± 0.06
+3	−1.64 ± 0.04	0.470 ± 0.083	2.54 ± 0.03
+4	−0.549 ± 0.059	0.304 ± 0.064	−0.246 ± 0.060
+5	−0.388 ± 0.081	−0.862 ± 0.106	−0.402 ± 0.034

The k_{int} of the wild type, R342A, and R342Y were 10.5 s^{-1}, 7.78 s^{-1}, and 1.31 s^{-1}, respectively.



Graphical abstract. Horikoshi, et al.

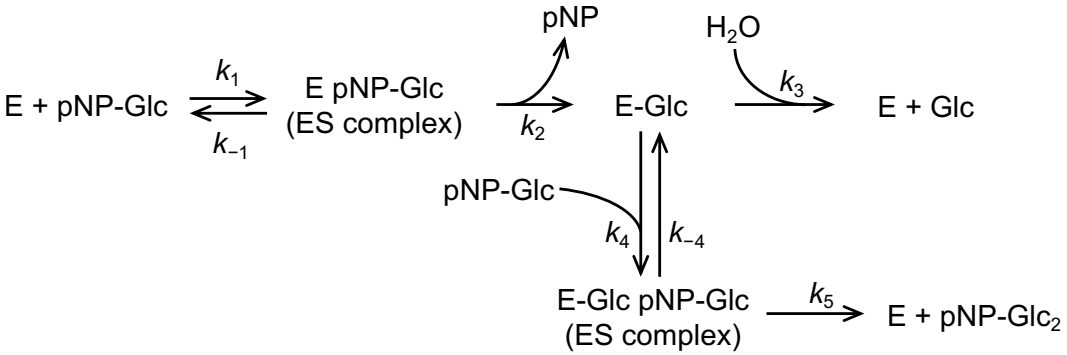


Figure 1. Horikoshi, et al.

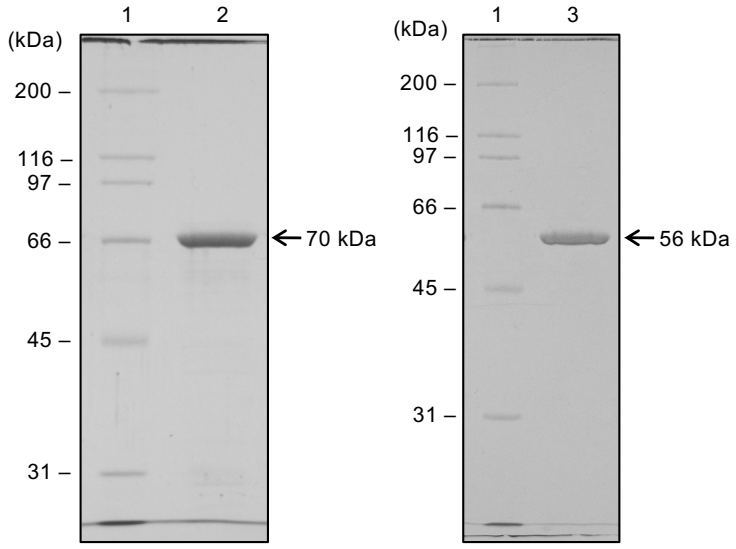


Figure 2. Horikoshi, et al.

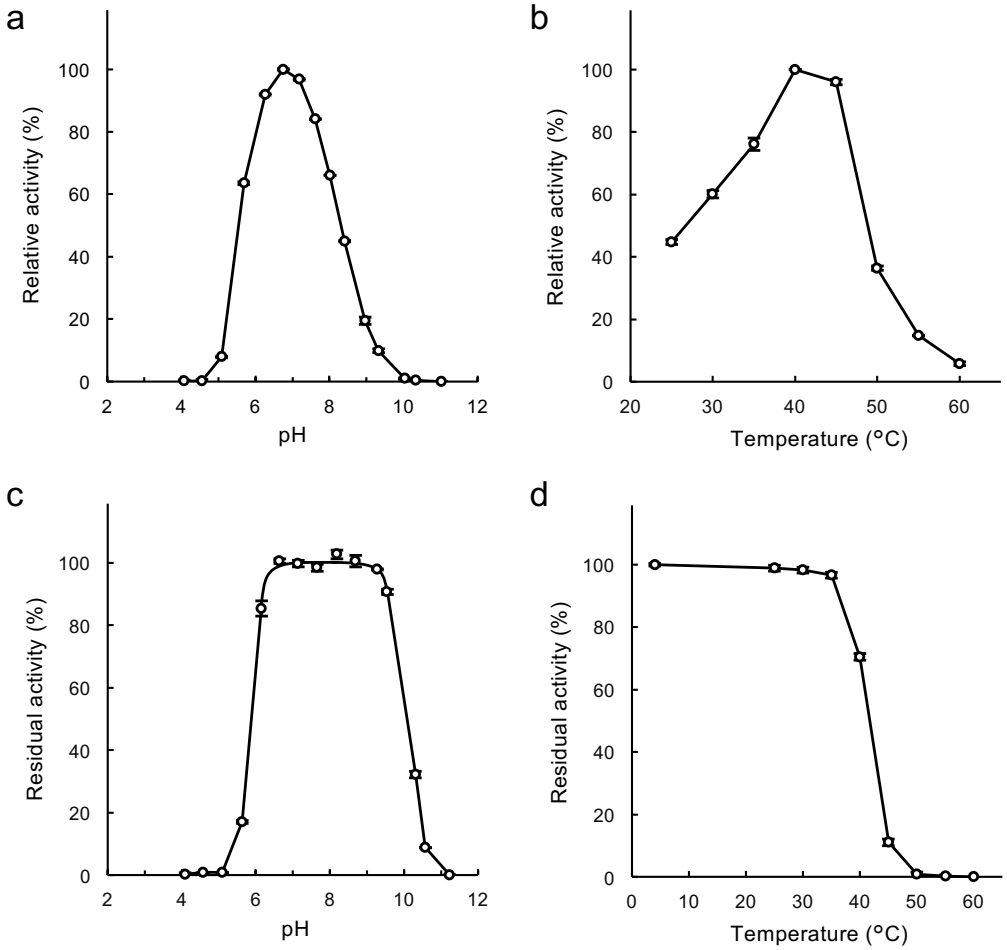


Figure 3. Horikoshi, et al.

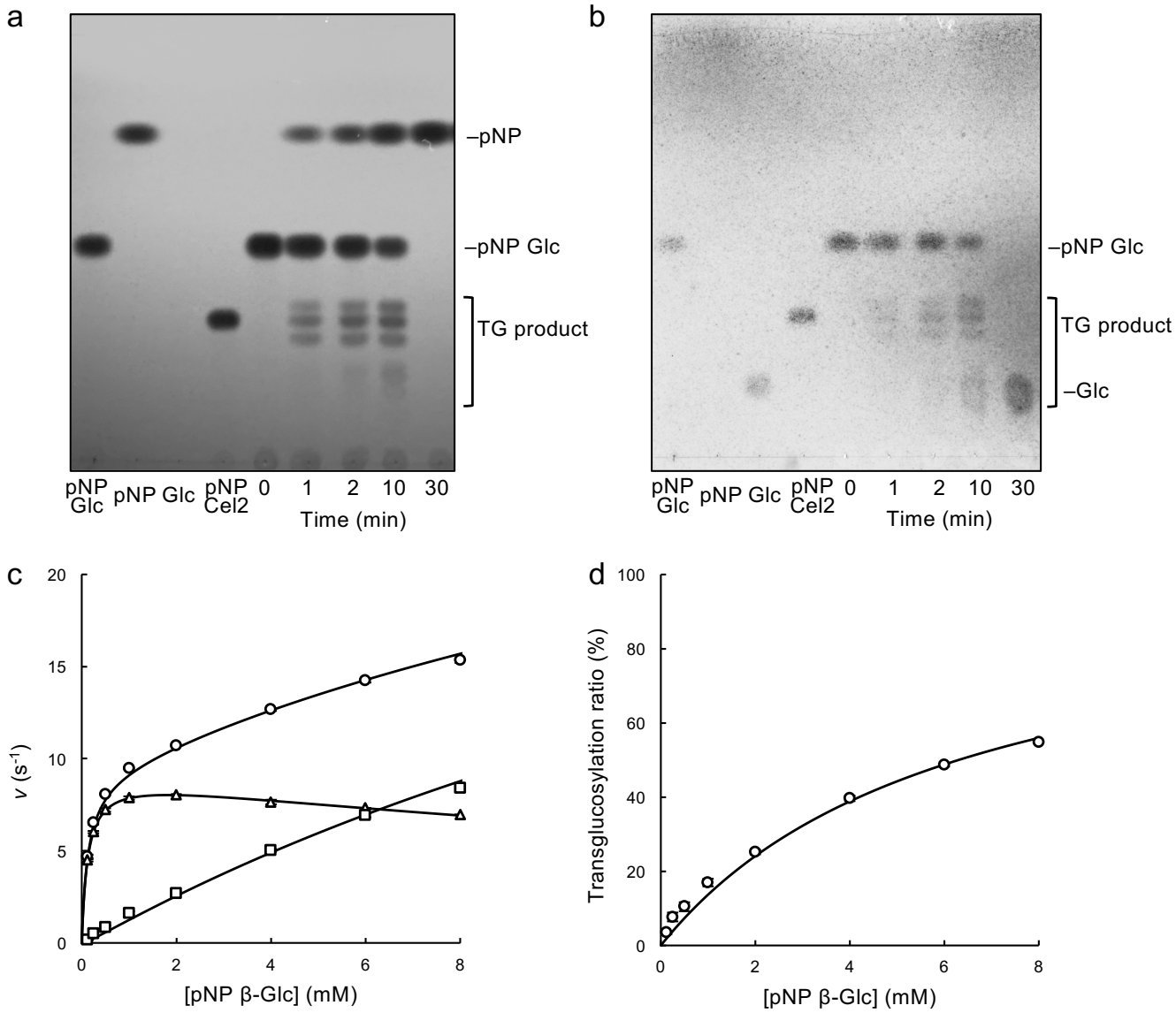


Figure 4. Horikoshi, et al.

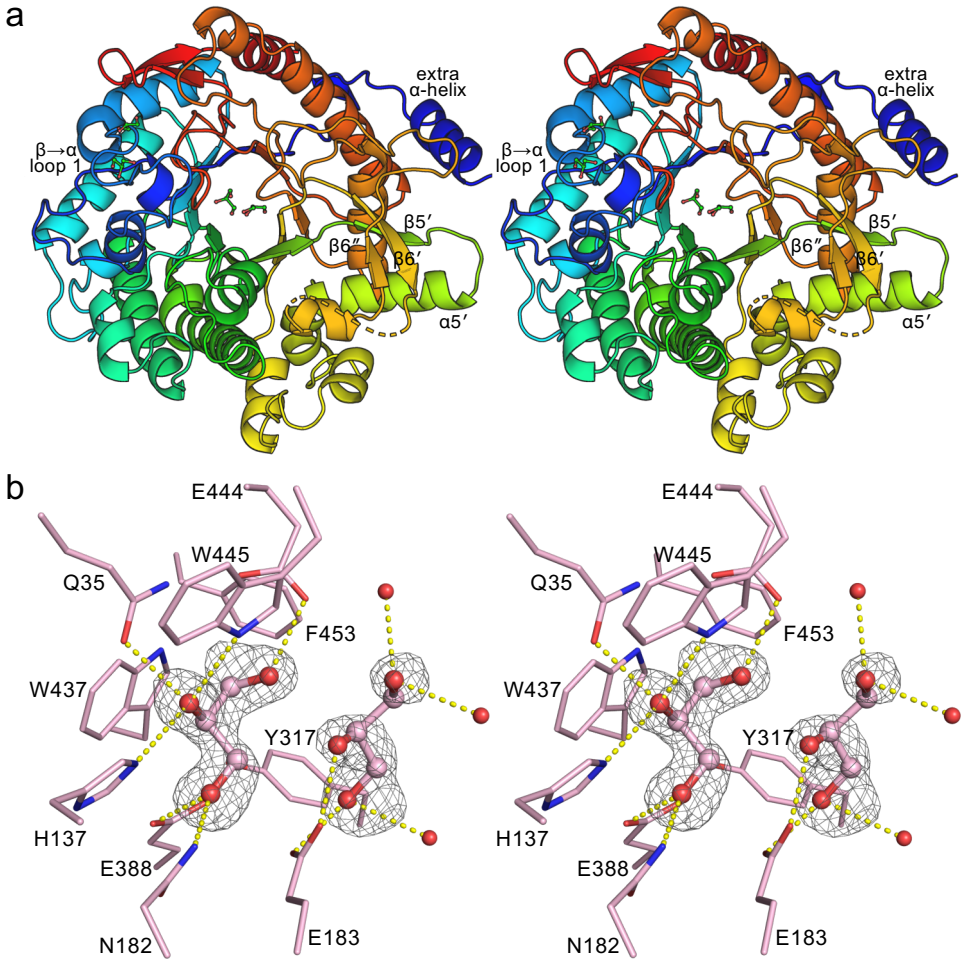


Figure 5. Horikoshi, et al.

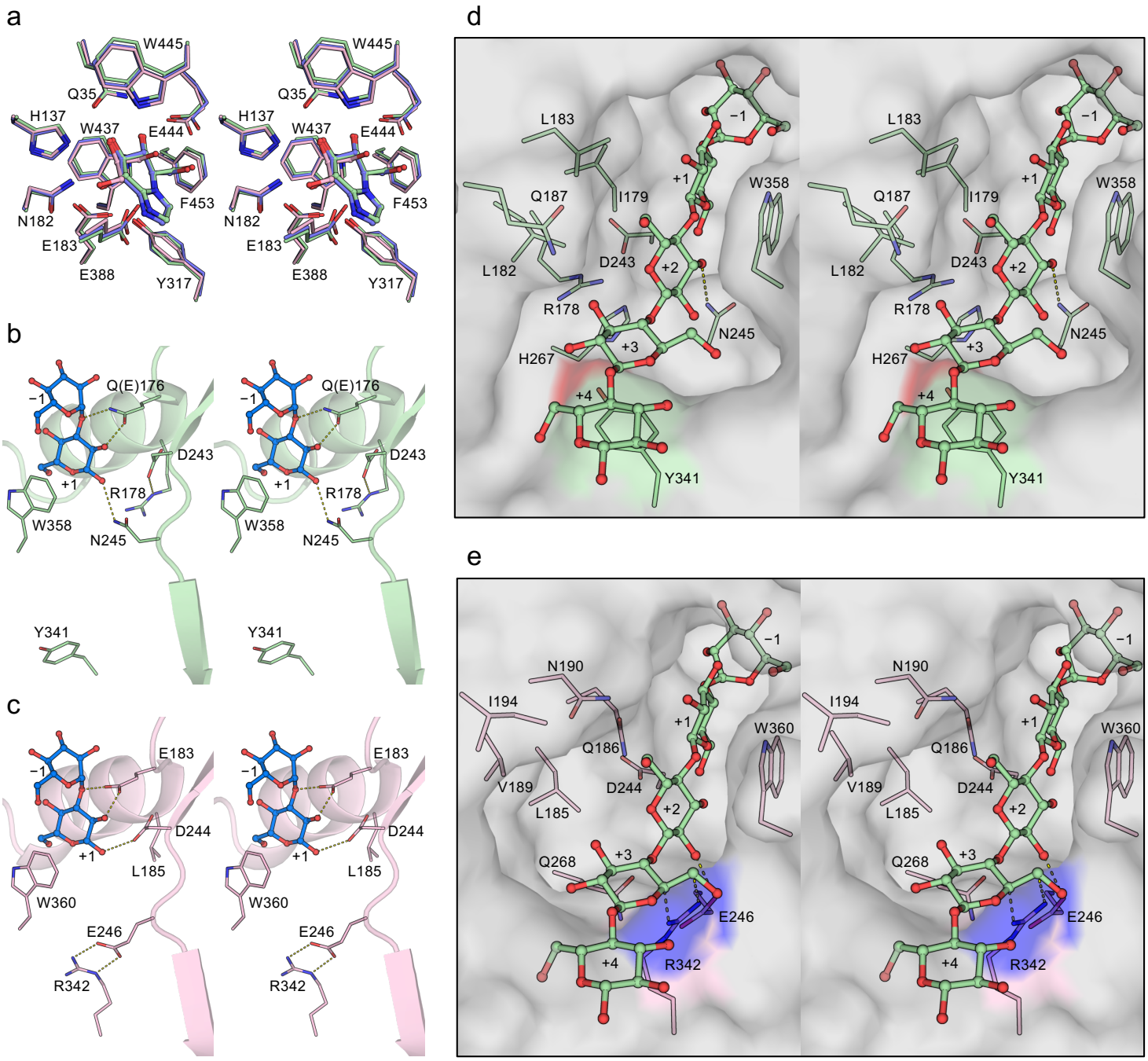


Figure 6. Horikoshi, et al.

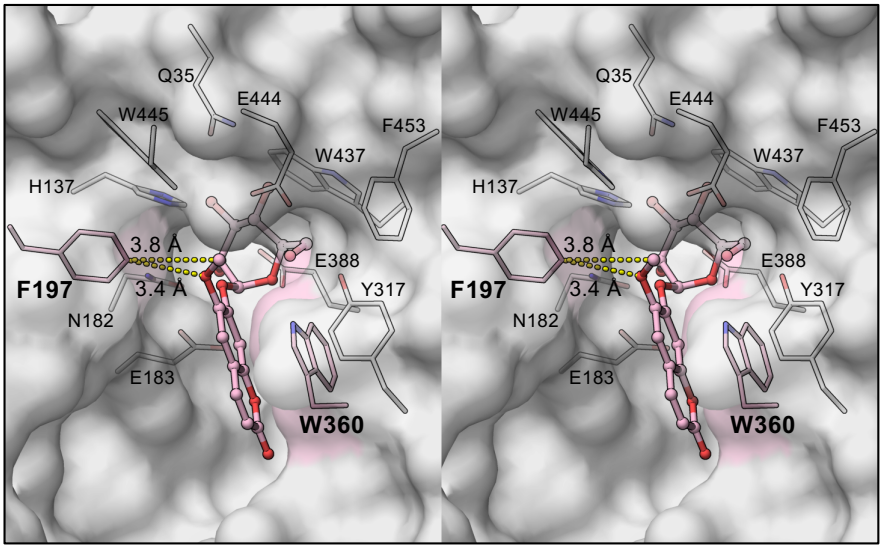


Figure 7. Horikoshi, et al.

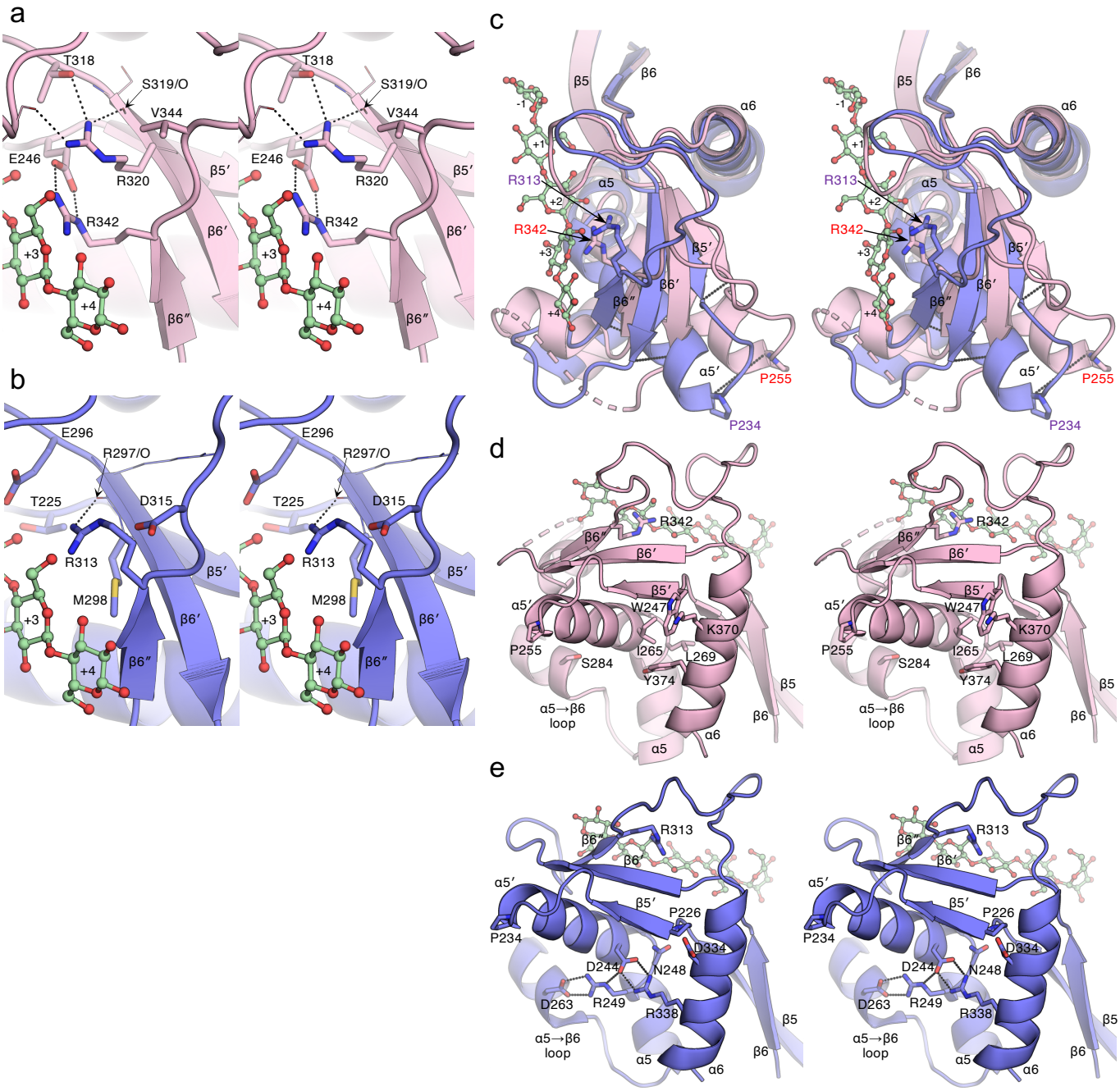


Figure 8. Horikoshi, et al.