Rational Design of a β-Glycosidase with High Regiospecificity for Triterpenoid Tailoring


Introduction

Many glycosylated natural products have been of great significance as promising sources of pharmaceuticals.[1] Triterpenoid saponins, which are glycosylated forms of triterpenoids produced by Panax ginseng, have recently emerged as potential therapeutics for cancer,[2] diabetes,[3] oxidative stress,[4] and aging.[5] The most abundant triterpenoid saponins are the protopanaxadiol (PPD) type, and usually exist as Rb1, Rb2, Rb3, and Rc in nature (Scheme 1).[6] These major triterpenoids have two sugar moieties, at C3 and C20 of the dammarane ring with glycosidic linkages, with four different kinds of sugar as the outer moiety at C20. Triterpenoids with the desired glycosylation patterns have been shown to attract considerable attention as potential therapeutics for inflammatory diseases and various types of cancer. Sugar-hydrolyzing enzymes with high substrate specificity would be far more efficient than other methods for the synthesis of such specialty triterpenoids, but they are yet to be developed. Here we present a strategy to rationally design a β-glycosidase with high regiospecificity for triterpenoids. A β-glycosidase with broad substrate specificity was isolated, and its crystal structure was determined at 2.0 Å resolution. Based on the product profiles and substrate docking simulations, we modeled the substrate binding modes of the enzyme. From the model, the substrate binding cleft of the enzyme was redesigned in a manner that preferentially hydrolyzes glycans at specific glycosylation sites of triterpenoids. The designed mutants were shown to produce a variety of specialty triterpenoids with high purity.

Triterpenoids with desired glycosylation patterns have attracted considerable attention as potential therapeutics for inflammatory diseases and various types of cancer. Sugar-hydrolyzing enzymes with high substrate specificity would be far more efficient than other methods for the synthesis of such specialty triterpenoids, but they are yet to be developed. Here we present a strategy to rationally design a β-glycosidase with high regiospecificity for triterpenoids. A β-glycosidase with broad substrate specificity was isolated, and its crystal structure was determined at 2.0 Å resolution. Based on the product profiles and substrate docking simulations, we modeled the substrate binding modes of the enzyme. From the model, the substrate binding cleft of the enzyme was redesigned in a manner that preferentially hydrolyzes glycans at specific glycosylation sites of triterpenoids. The designed mutants were shown to produce a variety of specialty triterpenoids with high purity.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201500004.

Department of Biological Sciences
Korea Advanced Institute of Science and Technology (KAIST)
Daehak-ro, Yuseong-gu, Daejeon, 305-701 (Korea)
E-mail: hskim76@kaist.ac.kr

[b] Dr. S.-G. Kim
Microbial Resource Center
Korea Research Institute of Bioscience & Biotechnology
Daejeon 305-806 (Korea)

[*] These authors contributed equally to this work.

Scheme 1. Chemical structures of PPD-type triterpenoids and their derivatives according to the sugar moieties at the C3 and C20 positions of a dammarane ring.[12] Glc, β-D-glucopyranosyl; arap, α-L-arabinopyranosyl; xyl, β-D-xylpyranosyl; ara, α-L-arabinofuranosyl; gyp, gypenoside; H, hydrogen. C stands for compound.
ing cleft of the enzyme in a way that it preferentially cleaves glycans at specific glycosylation sites of triterpenoids.

**Results and Discussion**

We isolated a β-glycosidase with promiscuous activity in terms of position and moiety of glycans on the aglycon backbone of triterpenoids from a bacterium (Supporting Information). The strain was identified as *Microbacterium sp. Gsoil 167*, and the gene coding for β-glycosidase was cloned and expressed in *Escherichia coli*. The enzyme consists of 486 amino acids, and a homology analysis revealed that the enzyme belongs to β-glycosidase family 1 (GH_1 family).\(^{[18]}\) The purified enzyme showed broad deglycosylation activity for oNP and pNP sugars (Table S1 in the Supporting Information), with the preference pNP-glucose > pNP-arabinopyranose > oNP-glucose > pNP-xylose. We determined the catalytic activity of the purified enzyme for the major triterpenoids, Rb1, Rb2, Rb3, and Rc (Fig-
The major products were identified as Rd and F2 from Rb1 and Rb2, C-Mx1 and F2 from Rb3, and C-Mc1 from Rc. Although F2 was not produced from Rc because of a lack of deglycosylation activity for arabinofuranose, the results demonstrate that the enzyme is able to cleave the outer sugar moieties at both C3 and C20 of the major triterpenoids, thus confirming the broad specificity, even with natural substrates.

To further investigate the catalytic feature of wild-type β-glicosidase, we traced the products from the major triterpenoids by HPLC analysis, and further identified them by LC/MS (Figures 1A, S2 and S3). The enzyme was shown to hydrolyze the outer sugar moiety at C20 and the outer glucose at C3 of Rb1 and Rb2 in a sequential manner (Figure 1A). Accordingly, the enzyme converted both Rb1 and Rb2 via the following pathway: Rb1 and Rb2 → Rd → F2 (Figure 1B). In the case of Rb3, on the other hand, the enzyme first hydrolyzed the outer glucose at C3, followed by cleavage of xylose at C20. The enzyme hydrolyzed only the outer glucose at C3 of Rc (Figure 1A); thus, the conversion of Rb3 and Rc by the wild-type enzyme is likely to proceed through two different pathways: Rb3 → C-Mx1 → F2 and Rc → C-Mc1 (Figure 1B). For all major triterpenoids, however, the enzyme showed negligble hydrolyzing activity for the inner glucose moieties at both C3 and C20. It is therefore evident that the enzyme is capable of eliminating only the outer sugar moieties at both C3 and C20 of the major triterpenoids, consistent with the preference shown for oNP- and pNP-sugars. It is also interesting that the outer sugar moieties of each position were cleaved in a highly sequential manner, according to the sugar position: C20 → C3 in both Rb1 and Rb2, or C3 → C20 in Rb3. Given that the enzyme follows only the preference for a sugar moiety, GypXVII and C-O would be simultaneously produced together with Rd, according to the ratio of relative enzyme activity for the respective sugars in Rb1 and Rb2 or Rd as well as C-Mx1 in Rb3. Based on the results, the wild-type enzyme is likely to have a preference for the glycan position as well as for the glycan itself.

To obtain some insight into the substrate binding site of the wild-type enzyme, we determined its crystal structure at 2.0 Å resolution (Table S2). The enzyme shows the typical TIM barrel structure of a β-glicosidase [18] (GH family 1), with a highly conserved active center (Glu172 and Glu342) and a sugar binding pocket (Figures S4 and S5). In contrast, the substrate binding cleft of the enzyme is quite different in shape from those of the other GH1 family enzymes (Figure S6). The N terminus of helix χε points towards i5, whereas the same helix of the other family enzymes is directed toward χε3 (Figure S7). As the helix points towards i5, the loop between χε and χδ, and the space filled by the same helix and anil in the other family enzymes is expanded to the substrate binding cleft. Consequently, the entrance of the substrate binding cleft is long and flat compared to those of the other family enzymes (Figures 2A and S6). Similarly to other family enzymes, the inside of the cleft is composed of hydrophobic residues. [20] Taken together, the enzyme is likely to stabilize the triterpenoids through contact with both sides of the PPD skeleton in the binding cleft. A cross-section view of the substrate binding cleft of the enzyme reveals an inverted clover-like shape: an active center and a sugar binding pocket at the bottom of the leaf (Figure 2B). The depth of the cleft is estimated at 16.7 Å from the entrance to the bottom. The entrance of the substrate binding cleft is tapered in the center by Ile184, thus effectively dividing the entrance into two sections of 13.2 and 11.6 Å (Figure 2A). The tapered region is only 3.4 Å wide, and would seems to hamper access of the PPD skeleton to this region. Thus, the substrates are presumed to bind to the enzyme through two different paths, P1 and P2 (Figure 2B).

Given the depth and the size at the two entrances of the substrate binding cleft, the major triterpenoids should enter longitudinally into the enzyme (Figure 2B) in order for the sugar moieties at C3 and C20 to reach the sugar binding pocket. As the structures of the triterpenoids differ at positions C3 and C20 by an additional alkene chain, we reasoned that each path of the enzyme has a substrate preference. To investigate the substrate preference with respect to glycan position, we attempted to co-crystallize substrates with catalytically inactive mutants, but failed. Therefore, we performed docking simulations with Rg3 and GypLXXV as representative substrates with a sugar moiety at C3 and C20. As expected, both substrates favored longitudinal access to the enzyme (Figure 2C and D). Interestingly, Rg3 was shown to bind the enzyme through P1 in multiple binding modes (Figure 2C), whereas GypLXXV bound through both paths (Figure 2D). This seems to be attributable to structural rigidity of a dammarane ring,
which is directly bonded to a sugar moiety at C3. Because P1 is relatively straight and wide compared to P2 (Figure 2B), P1 might be preferred over P2 in the binding of a sugar at C3.

According to the modeled substrate binding modes of the wild-type enzyme, the sugars at both positions of the triterpenoid can access the active site of the enzyme through P1. In contrast, P2 will favor the binding of the sugar at C20. Thus, the P1-perturbed enzyme will specifically cleave the sugar moiety at C20. Likewise, the P2-perturbed enzyme is expected to hydrolyze the sugars at C3 as well as C20. Based on the structural analysis of the substrate binding cleft, we identified two residues that appear to restrict substrate access to each path. Glu316 is in the bottleneck of P1, and Gln216 is in the right leaf of the substrate binding cleft of P2 (Figure 2C and D). We changed both residues to Trp to construct mutants E316W and Q216W (Figure 3A), and analyzed the reaction products from four major triterpenoids by the mutants by using TLC and LC/MS (Figures 3B, S3 and S7). As expected, mutant E316W specifically cleaved the sugar moiety at C20 to produce Rd from Rb3 (Figure 3B, lane 13, Figures 3C and S8), whereas the wild-type enzyme produced C-Mx1 (Figure 3B, lane 12, and Figure S8). The catalytic efficiency ($k_{cat}/K_m$) of E316W was about four-fold higher than that of wild-type in the conversion of Rb1 to Rd (Table 1). These results indicate the preference of E316W was shifted to a sugar at C20; the production of Rd from Rb1 and Rb2 significantly increased compared to the wild-type enzyme (Figure 3B, lanes 4 and 8). In the case of the P2-perturbed mutant (Q216W), random cleavage of the sugars at both positions was expected. However, Q216W predominantly cleaved the outer glucose at C3 of Rb1 and Rb2 to produce GypXVII and C-O respectively, as the first intermediates (Figure 3B, lanes 5 and 9, Figures 3C and S8). For Q216W, $k_{cat}/K_m$ was about seven times higher than that of the wild-type enzyme (Table 1). This indicates that the preference of the mutant shifted toward a sugar at C3, probably a consequence of the difference in binding energy between the sugars at C3 and C20. The alkene chain at C20 should be in the binding cleft when the sugar moiety at C20 binds in the active site (Figure 2D). Because of the flexibility of this alkene chain, the binding of the sugar moiety at C20 will be relatively unfavorable relative to that at C3; this led the Q216W mutant to predominantly cleave the outer glucose at C3. Consequently, the production of C-Mx1 and C-Mc1 from Rb3 and Rc by Q216W increased significantly compared to the wild-type (Figure 3B, lanes 14 and 19). Collectively, these results demonstrate that regiospecificity for triterpenoids can be rationally
In order for the inner sugar moiety of a substrate to be cleaved, deeper access of a substrate into the binding cleft is necessary. We therefore attempted to enlarge the entrance of the substrate binding cleft of the wild-type enzyme, by mutating Ile184, Ile389, and Phe390, which are located at the tapered region of the substrate entrance, to Ala (mutant 3MT). As a result, the entrance expanded to 7 Å at the tapered region, thus enabling access of the PPD skeleton (Figures 4A and S9). The resulting mutant was subjected to reactions with four major triterpenoids, and the product profiles were analyzed by HPLC and LC/MS (Figures 4B and S3). 3MT primarily

Table 1. Kinetic constants of wild-type and mutant enzymes for various substrates.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>Product</th>
<th>$K_m$ [mM]</th>
<th>$k_{cat}$ [min$^{-1}$]</th>
<th>$k_{cat}/K_m$ [mM$^{-1}$ min$^{-1}$]</th>
<th>Fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>Rb1</td>
<td>Rd</td>
<td>1.9 ± 0.5</td>
<td>6.4 ± 2.5</td>
<td>4.0 ± 2.4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>C-K</td>
<td>n.d.</td>
<td>n.d.</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>E316W</td>
<td>Rb1</td>
<td>Rd</td>
<td>1.7 ± 0.3</td>
<td>25.4 ± 3.3</td>
<td>15.8 ± 4.7</td>
<td>4.0</td>
</tr>
<tr>
<td>Q216W</td>
<td>Rb1</td>
<td>GypXVII</td>
<td>0.9 ± 0.1</td>
<td>24.5 ± 1.4</td>
<td>27.7 ± 4.6</td>
<td>7.0</td>
</tr>
<tr>
<td>3MT</td>
<td>F2</td>
<td>C-K</td>
<td>0.7 ± 0.2</td>
<td>0.1 ± 0.0</td>
<td>0.2 ± 0.0</td>
<td>--</td>
</tr>
</tbody>
</table>

[a] n.d.: not determined.

Figure 4. A) Substrate binding mode of the 3MT (I184A/I389A/F390A) mutant. B) Product profiles of major triterpenoids produced by 3MT. Details of STD are shown in Figure S2. C) Biotransformation pathways of the major triterpenoids by 3MT. The arab of Rb2 and C-O, and arab of Rc, C-Mc1, and C-Mc (corresponding to the outer sugar-moieties at C20) are shown in parenthesis; connectivity is denoted by asterisks.
cleaved the outer glucose at C3 of the major triterpenoids (Figure 4C) as did the Q216W mutant (Figure 3C). This regiospecificity is likely to stem from the unfavorable binding of the glucose at C20 because of the flexible alkenic chain at C20. Then the outer sugars at C20 were cleaved in Rb1 and Rb2 (Figure 4C), followed by cleavage of the inner glucose at C3. In the case of Rb3 and Rc, the inner glucose at C3 was then hydrolyzed, and finally xylose at C20 of Rb3 was cleaved (Figure 4C). The production rates of C-K through cleavage of inner glucose at C3 of Rb1, Rb2, and Rb3 were estimated at 0.3, 0.4, and 0.2 μM min⁻¹, respectively. The production rates of C-Mc and C-Mc were estimated at 0.3 and 1.0 μM min⁻¹, respectively. In contrast, production of C-K by the wild-type enzyme was negligible. Thus, the conversion of Rb1 (or Rb2) and Rb3 (or Rc) by 3MT is likely to proceed through two different reaction pathways: Rb1 (Rb2) → GypXVII (C-O) → F2 → C-K and Rb3 (Rc) → C-Mc1 (C-Mc1) → C-Mc → C-K (Figure 4C). The results indicate that 3MT can cleave the inner glucose at C3 of a substrate because the entrance to the substrate binding cleft is enlarged. 3MT produced F2 from Rb1 with 89% purity, C-K from Rb1 and Rb2 with 61 and 71% purities, respectively, C-Mc1 from Rb3 with 100% purity, and C-Mc1 and C-Mc from Rc with 100% purity (Table 2 and Figure S10). Cleavage of the inner glucose at C20 was not observed for any of the major triterpenoids. This might be because of steric hindrance between the alkenic chain of the PPD skeleton and 3MT.

Conclusions

We have demonstrated a strategy to design a β-glycosidase for producing specialty triterpenoids with high regiospecificity, based on product profiling and docking simulation. Enzymatic methods to produce specialty triterpenoids have mainly focused on the isolation of sugar-hydrolyzing enzymes, but these approaches resulted in limited types of triterpenoids, mainly owing to undesirable substrate specificity of enzymes. Starting from a promiscuous β-glycosidase, we were able to redesign the active site of the enzyme in a manner that specifically hydrolyzes the glycans at specific glycosylation sites of triterpenoids. The designed mutants were shown to produce a variety of specialty triterpenoids with high purity, thus validating this approach. Our strategy can be applied to other sugar-hydrolyzing enzymes showing catalytic activity toward diverse triterpenoid saponins, such as protopanaxatriol (PPT) and oleanane types.

Acknowledgements

This work was supported by the National Research Foundation (NRF) grant (2014-011498) from the Ministry of Science, ICT and Future Planning and Brain Korea 21 funded by the Ministry of Education.

Keywords: binding mode · glycosides · rational design · regiospecificity · terpenoids

Table 2. Reaction products from major triterpenoids by wild-type β-glycosidase and three mutants.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Reaction products[a]</th>
<th>WT</th>
<th>E316W</th>
<th>Q216W</th>
<th>3MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rb1 Rd</td>
<td>F2 (79) Rd</td>
<td>GypXVII Rd</td>
<td>GypXVII (39, F2 (89), C-K (61))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rb2 Rd</td>
<td>F2 (74) Rd</td>
<td>C-O Rd</td>
<td>C-O (67, F2 (61), C-K (71))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rb3 C-Mc1 F2 Rd</td>
<td>C-Mc1 (100), C-Mc (40), C-K (47)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rc C-Mc1</td>
<td>C-Mc1 C-Mc1 C-Mc1 (100), C-Mc (100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[a] Values in parenthesis are relative proportion (%) of specialty triterpenoids at the times specified in Figures S10 and S11.


[21] The X-ray crystal structure of wild-type BGL167 has been deposited at the Protein Data Bank under accession code 4R27.

Received: January 5, 2015
Published online on February 20, 2015