










C-Glycoside metabolism in the gut and in nature: Identification, characterization, structural analyses and distribution of C-C bond-cleaving enzymes

Takahiro Mori ^{1,2,3,8}, Takuto Kumano ^{4,5,8}, Haibing He^{1,8}, Satomi Watanabe^{4,8}, Miki Senda ⁶, Toshio Moriya ⁶, Naruhiko Adachi ⁶, Sanae Hori⁴, Yuzu Terashita⁴, Masato Kawasaki⁶, Yoshiteru Hashimoto^{4,5}, Takayoshi Awakawa ^{1,2}, Toshiya Senda ^{6,7}✉, Ikuro Abe ^{1,2}✉ & Michihiko Kobayashi ^{4,5}✉

C-Glycosides, in which a sugar moiety is linked via a carbon-carbon (C-C) bond to a non-sugar moiety (aglycone), are found in our food and medicine. The C-C bond is cleaved by intestinal microbes and the resulting aglycones exert various bioactivities. Although the enzymes responsible for the reactions have been identified, their catalytic mechanisms and the generality of the reactions in nature remain to be explored. Here, we present the identification and structural basis for the activation of xenobiotic C-glycosides by heterocomplex C-deglycosylation enzymes from intestinal and soil bacteria. They are found to be metal-dependent enzymes exhibiting broad substrate specificity toward C-glycosides. X-ray crystallographic and cryo-electron microscopic analyses, as well as structure-based mutagenesis, reveal the structural details of these enzymes and the detailed catalytic mechanisms of their remarkable C-C bond cleavage reactions. Furthermore, bioinformatic and biochemical analyses suggest that the C-deglycosylation enzymes are widely distributed in the gut, soil, and marine bacteria.

¹Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ²Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan. ³PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. ⁴Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan. ⁵Microbiology Research Center for Sustainability, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan. ⁶Structural Biology Research Center, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan. ⁷Faculty of Pure and Applied Sciences, University of Tsukuba, 1-1-1, Tennodai, Ibaraki 305-8571, Japan. ⁸These authors contributed equally: Takahiro Mori, Takuto Kumano, Haibing He, Satomi Watanabe. ✉email: toshiya.senda@kek.jp; abei@mol.f.u-tokyo.ac.jp; kobayashi.m.fe@u.tsukuba.ac.jp

The human gastrointestinal tract contains a complex microbial ecosystem that plays a pivotal role in human health through the maintenance of metabolic homeostasis, regulation of host immunity, and defense against pathogens¹. In addition to those functions, the transformation of orally ingested xenobiotic chemicals by intestinal microbes is also important for the activation of prodrugs or the inactivation of toxic compounds. However, the molecular mechanisms underlying the activation of bioactive compounds by enzymes in intestinal bacteria remain largely unknown.

Deglycosylation by intestinal microbes is a crucial reaction for the absorption and/or exertion of the biological activities of various compounds^{2–4} that we ingest in food and drugs, as exemplified by the reactivation of drug glucuronide metabolites and the absorption of glycosylated plant-derived bioactive compounds^{5–13}. Among several deglycosylation pathways, we focused on the pathway metabolizing C-glycosides, of which over 300 kinds have been identified in nature¹⁴. For example, puerarin and mangiferin have been isolated from the roots of *Pueraria lobata* (“Kudzu”), which is used in traditional Japanese herbal medicine, and from mango peel, respectively. Carminic acid, a C-glycoside extracted from cochineal beetles, is used for food staining as a natural “red dye”, over 200 tons being used per year worldwide. Compared with other glycosides (O-, N- and S-glycosides), C-glycosides are quite stable against chemical and enzymatic treatments¹⁵.

Over the past few decades, several C-C bond cleavage reactions for plant C-glycosides due to intestinal bacteria have been reported^{5–11,16}. Two metabolic gene clusters, the *dfg* and *dgp* clusters, responsible for the C-deglycosylation of flavonoid C-glycosides, were recently found in human intestinal bacteria^{6,16,17}. DfgA-E in *Eubacterium cellulosolvens* catalyze the deglycosylation of homoorientin and isovitexin⁶, while DgpA (Gfo/Idh/MocA family oxidoreductase), DgpC (α -subunit; sugar isomerase/epimerase-like enzyme), and DgpB (β -subunit; hypothetical protein) in the *dgp* cluster from a PUE bacterial strain catalyze the two-step C-C bond cleavage of puerarin¹⁸, which shows various pharmacological properties such as antioxidant, anticancer and anti-inflammatory activities, and attenuation of insulin resistance¹⁹ (Fig. 1a). DgpA catalyzes the oxidation of the glucoside moiety of puerarin to generate 3'-oxo-puerarin, in which the 3'-oxo-sugar moiety is cleaved by an unusual C-deglycosylation enzyme complex, which is an $\alpha 4\beta 4$ heterooctamer of DgpB-C, to produce daidzein. DgpC shares less than 20% amino acid sequence similarity with other sugar isomerases^{18,20–22}, which do not form a heterocomplex to exert their activities. Although the C-C bond cleavage reactions due to intestinal bacteria are attracting keen interest, the detailed physicochemical properties of the enzymes, their crystal structures, catalytic residues, and catalytic mechanisms, and the generality of the reactions in nature remain unclear.

Here, we describe the X-ray and cryo-EM structures as well as the identification and characterization of C-deglycosylation enzymes from intestinal and soil bacteria, and the structural basis of the common C-deglycosylation reaction by the unique enzymes. The generality of the C-glycoside-metabolisms in the gut, soil, and marine bacteria was investigated by biochemical and bioinformatic analyses.

Results

Phylogenetic analysis of C-glycoside deglycosidase (CGD) enzymes. A database search using the amino acid sequences of DgpB and DgpC as queries indicated that both aerobic and anaerobic bacteria, phyla Actinobacteria, Proteobacteria, and Firmicutes, from soil and marine environments and the human

digestive system, have each of DgpB and DgpC homologs (Supplementary Fig. 1). As far as we investigated, their gene homologs existed side by side in each bacterial genome (Supplementary Fig. 2 and Supplementary Table 1), suggesting that they form a complex such as DgpB-C to exhibit C-glycoside deglycosidase (CGD) activity. Phylogenetic analysis of the homologs revealed that the CGDs share 40–50% amino acid sequence similarity with one another even in different microbial phyla.

Characterization of intestinal C-deglycosylation enzymes. A previous study demonstrated that the combination of five Dfg enzymes, DfgA-E, from *E. cellulosolvens* catalyzed the deglycosylation of the flavone C-glycosides homoorientin and isovitexin⁶. Here, sequence analysis revealed that β -galactosidase DfgA and hypothetical protein DfgB share sequence similarity with DgpC and DgpB, respectively (34% and 38% amino acid identities), suggesting that the pair of DfgA and DfgB is involved in the cleavage of the C-C bond (Supplementary Table 1). Expectedly, our studies, such as co-expression, purification, and SEC-MALS analysis of DfgA and DfgB, revealed that DfgA-B formed an $\alpha 4\beta 4$ heterooctamer in solution, as in the case of DgpB-C (Supplementary Fig. 3a and Supplementary Table 2).

To characterize the biochemical properties of the enzyme complexes in C-deglycosylation reactions, the purified DgpB-C and DfgA-B proteins were subjected to in vitro analyses. For clarity, the DgpB-C and DfgA-B complexes are referred to as *PuCGD* and *EuCGD*, respectively. First, the metal ion dependency of *PuCGD* and *EuCGD* was investigated by the addition of each metal to the chelate reagent-treated enzymes, using 3'-oxo-puerarin or 3'-oxo-homoorientin as a substrate. While the chelate reagent-treated *PuCGD* and *EuCGD* showed activity reduced to 16% and 37% as compared to the non-chelate reagent-treated enzyme (represented as WT in Supplementary Fig. 4a), respectively, the activity was restored by the addition of divalent metal ions other than Cu^{2+} and Fe^{2+} . The addition of Ni^{2+} and Mn^{2+} to each of *PuCGD* and *EuCGD* improved their activities up to 157% and 152% (for *PuCGD*) and 131% and 112% (for *EuCGD*), respectively. (Supplementary Fig. 4). These findings indicated that each of the divalent metal ions facilitated the C-C bond cleavage reaction. The optimal pH and temperature of both *PuCGD* and *EuCGD* were pH 6.0 and 60 °C (Supplementary Fig. 4b and c). To elucidate the substrate specificity of these enzymes, each of 3'-oxo-C6- or C8-glycosylated flavonoids, 3'-oxo-carminic acid and 3'-oxo-mangiferin **1–6**, was incubated with *PuCGD* and *EuCGD* (Figs. 1, 2, and Supplementary Fig. 5). As a result, *PuCGD* accepted the C8-glycosylated compounds **1** and **2**, but the C6-glycosylated compounds **3–6** were inert substrates. In contrast, *EuCGD* acted on C6-glycosylated compounds **3–5**, but not on C8-glycosylated compounds **1** and **2**. The steady-state kinetics values for deglycosylation reactions by *PuCGD* were $K_M = 12$ mM and $k_{\text{cat}} = 0.41$ /min for **1** and $K_M = 0.79$ mM and $k_{\text{cat}} = 0.020 \times 10^{-3}$ /min for **2**. On the other hand, the steady-state kinetics values for deglycosylation reactions by *EuCGD* were $K_M = 11$ mM and $k_{\text{cat}} = 0.27$ /min for **3** and $K_M = 3.6$ mM and $k_{\text{cat}} = 0.18$ /min for **4** (Fig. 2).

Identification of the C-deglycosylation enzyme from soil bacteria. As described above, phylogenetic analysis of and a database search for *PuCGD* and homologs suggested wide distribution of CGDs in nature. Indeed, in the genome of *Microbacterium* sp. 5-2b, which we recently discovered as a C-glycoside carminic acid-metabolizing bacteria on assimilation screening of a soil environment, gene products sharing 29/49% and 30/52% amino acid identity/similarity with the α - and β -subunits of *PuCGD* were found and designated as CarB and CarC, respectively

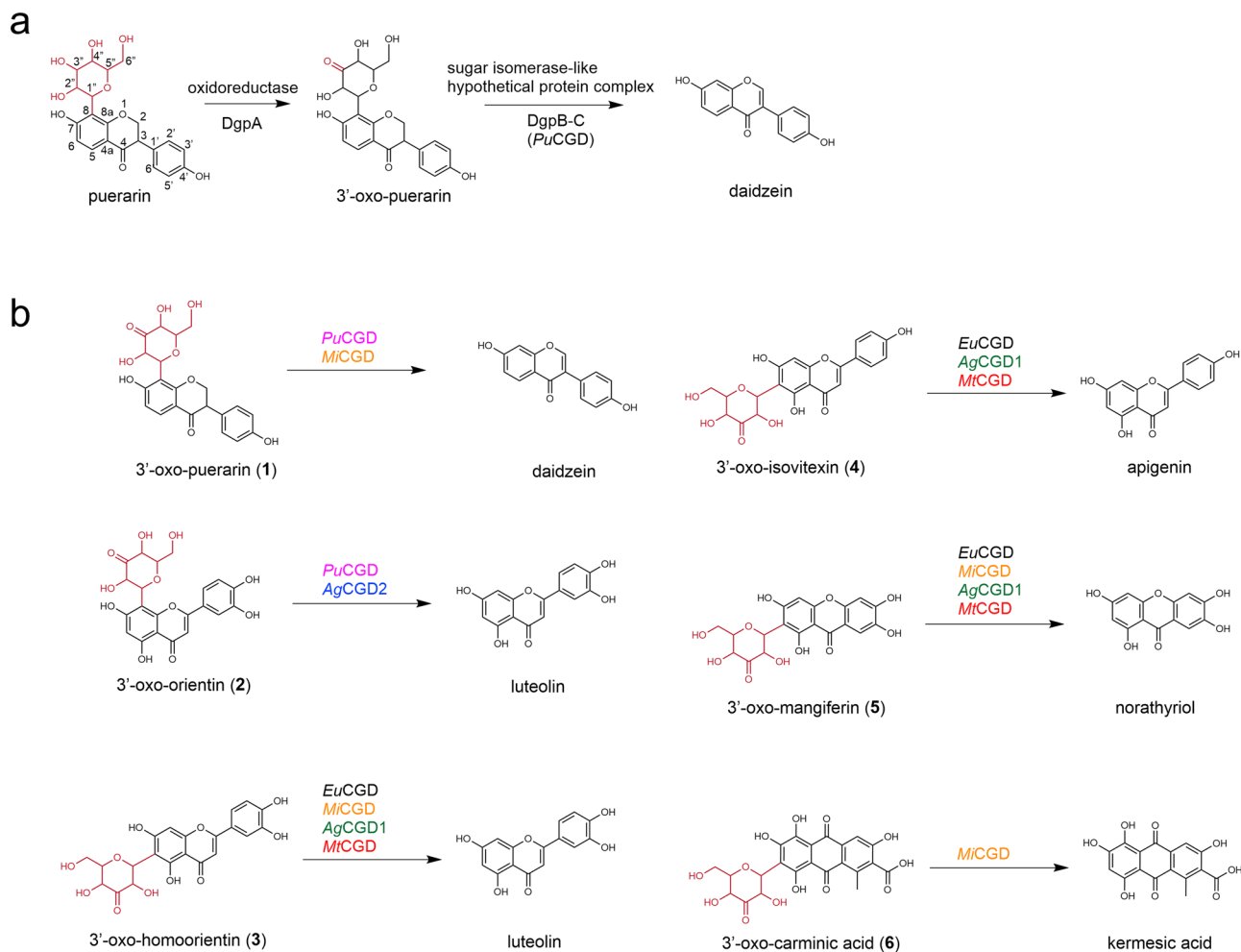


Fig. 1 Reactions of C-deglycosylation enzymes. **a** Two-step C-C bond cleavage reaction by oxidoreductase (DgpA) and C-deglycosylation enzymes (DgpC and DgpB; PuCGD). **b** C-deglycosylation enzymes used in our studies: PuCGD, EuCGD, MiCGD, AgCGD1, AgCGD2, and MtCGD. Sugar moieties are shown in red. The substrates investigated in this study were C8-glycosylated flavonoids **1** and **2**, C6-glycosylated flavonoids **3** and **4**, C-glycosylated xanthone **5**, and C-glycosylated anthraquinone **6**.

(Supplementary Fig. 2 and Supplementary Table 1). No other gene products in *Microbacterium* sp. 5-2b showed similarity to PuCGD. To examine the universality of C-deglycosylation reactions and their biochemical functions, CarB and CarC were expressed individually and incubated with 3'-oxo-carminic acid as a substrate (Supplementary Fig. 6). On incubation of CarB and CarC together, 3'-oxo-carminic acid was converted to kermesic acid, although neither CarB nor CarC showed 3'-oxo-carminic acid-converting activity (Supplementary Fig. 7). We next co-expressed His-tagged CarB and tag-free CarC. As a result, His-tagged CarB was purified together with CarC by using a Ni-affinity column (Supplementary Fig. 3b). SEC-MALS analysis revealed that the molecular mass of the CarB-C complex was 49.0 kDa. Together with the theoretical masses of CarB (39.4 kDa) and CarC (19.6 kDa) calculated from their deduced amino acid sequences, these findings indicated that CarB and CarC formed a heterodimer, while PuCGD and EuCGD formed an $\alpha_4\beta_4$ heterooctamer. We here designated CarB and CarC as the α -subunit and β -subunit of MiCGD, respectively. The purified MiCGD showed the catalytic activity converting 3'-oxo-carminic acid to kermesic acid without any cofactors as in the case of PuCGD and EuCGD (Fig. 1b). The absorption spectrum of the purified MiCGD showed an absorbance maximum near 280 nm. No other absorption peak or shoulder was observed at higher wavelengths

(Supplementary Fig. 6c). These results suggested that no cofactor which exhibits UV absorption was bound to the purified MiCGD enzyme.

A BLAST search identified more putative CGD homologs in soil bacteria. Two homologs from *Arthrobacter globiformis*, and one from *Microbacterium trichothecenolyticum* were designated as AgCGD1, AgCGD2, and MtCGD, respectively, and heterologously expressed in *E. coli* (Supplementary Figs. 1a, 2, and 3b, and Supplementary Table 1). The SEC-MALS analysis revealed that the purified enzymes of soil bacteria were $\alpha\beta$ heterodimers like MiCGD (Supplementary Table 2). The optimal temperatures of soil bacterial CGDs, including MiCGD, AgCGD1, AgCGD2, and MtCGD, were around 40 °C (Supplementary Fig. 4d). The optimal pH for the activities of MiCGD and AgCGD1 was pH 7.5, and that for MtCGD was pH 6.0 (Supplementary Fig. 4e).

Substrate specificity analysis of the soil bacteria-derived enzymes as to compounds **1-6** demonstrated that they also catalyzed the C-C bond cleavage of various C-glycosylated compounds (Figs. 1b, 2, and Supplementary Fig. 8). MiCGD showed high activity toward **6** and weakly accepted **1** and **3-5**. AgCGD1 and MtCGD showed activity toward **3-5**, while AgCGD2 only accepted **2** as a substrate. These results indicated that MiCGD, AgCGD1 and MtCGD mainly accepted the C6-glycosylated compounds as substrates, while AgCGD2 accepted

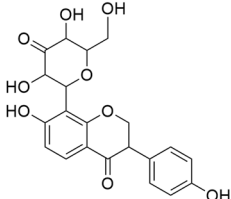
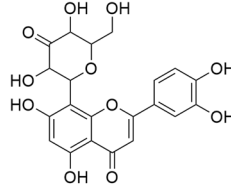
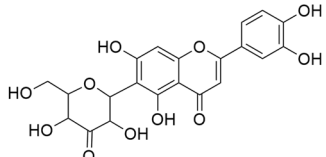
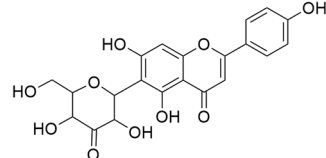
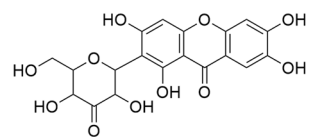
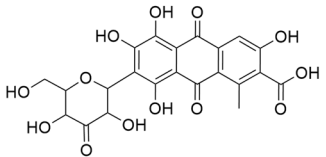
Substrate	<i>Pu</i> CGD	<i>Eu</i> CGD	<i>Mi</i> CGD	<i>Ag</i> CGD1	<i>Ag</i> CGD2	<i>Mt</i> CGD
	Deglycosylation activity [$\frac{K_M \text{ (mM)}}{k_{cat} \text{ (min}^{-1}\text{)}} \text{]}$					
 3'-oxo-puerarin (1)	12 ± 1 0.41 ± 0.02	N.D.	7.5 ± 2.5 0.12 ± 0.02	N.D.	N.D.	N.D.
 3'-oxo-orientin (2)	0.79 ± 0.07 0.000020 ± 0.000001	N.D.	N.D.	N.D.	> 10 40 ± 5	N.D.
 3'-oxo-homoorientin (3)	N.D.	11 ± 1 0.27 ± 0.01	11 ± 2 0.024 ± 0.003	3.4 ± 0.3 0.57 ± 0.02	N.D.	>10 21 ± 2
 3'-oxo-isovitexin (4)	N.D.	3.6 ± 0.4 0.18 ± 0.01	N.D.	3.5 ± 0.4 4.6 ± 0.2	N.D.	8.2 ± 1.8 32 ± 5
 3'-oxo-mangiferin (5)	N.D.	> 5 N.D.	5.4 ± 1.0 0.16 ± 0.02	3.7 ± 0.5 0.17 ± 0.01	N.D.	6.1 ± 1.4 0.60 ± 0.07
 3'-oxo-carminic acid (6)	N.D.	N.D.	0.39 ± 0.06 23 ± 1	N.D.	N.D.	N.D.

Fig. 2 Substrate specificities of C-deglycosylation reactions. The kinetic parameters of each C-deglycosylation enzyme towards C-glycosides were shown. In each set of numbers, the upper and lower numbers represent K_M (mM) and k_{cat} (min^{-1}), respectively. N.D.: not determined.

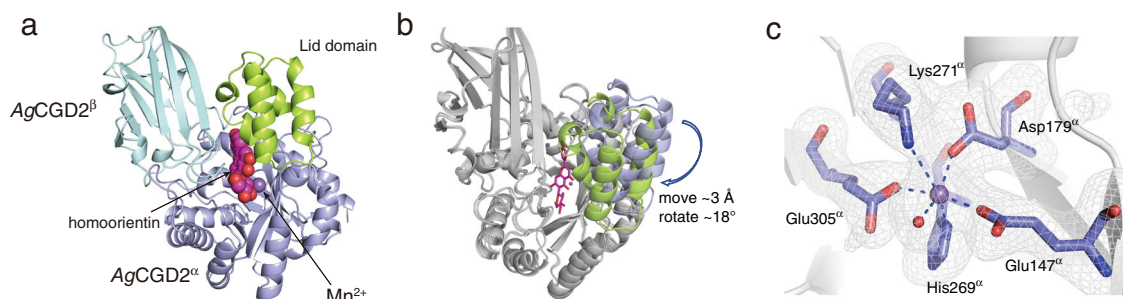


Fig. 3 Crystal structures of C-deglycosylation enzymes. **a** Heterodimer structure of AgCGD2. The β -sandwich structure of AgCGD2 $^{\beta}$ is shown in pale blue and the TIM barrel structure of AgCGD2 $^{\alpha}$ in purple. The lid domain of AgCGD2 $^{\alpha}$ is shown in light green. **b** Conformational change of the lid domain of AgCGD2 $^{\alpha}$ upon substrate binding. The lid domains in the apo structure and the structure complexed with homoorientin are shown in purple and light green, respectively. **c** The Fo-Fc polder omits the map of the metal-binding site of AgCGD2 $^{\alpha}$. The electron density map of the ligands is represented by a gray mesh, contoured at +3.0 sigma. Purple and red balls represent Mn $^{2+}$ and water, respectively.

the C8-glycosylated compound. The steady-state kinetics values for deglycosylation reactions by those enzymes were calculated (Fig. 2). The K_M value of MiCGD for **6** was 0.39 mM. On the other hand, most K_M values were around 3 to 10 mM.

Similar to the intestinal enzymes (*PuCGD* and *EuCGD*), after dialysis with chelating agents, the enzymatic activities of MiCGD, AgCGD1, and AgCGD2 were reduced to 13, 1.4, and 0%, respectively. As compared with the non-dialyzed enzyme (denoted as WT in Supplementary Fig. 4a), the activities of MiCGD and AgCGD1 were restored to 124% and 104%, respectively, by adding Mg $^{2+}$. On the other hand, the activity of AgCGD2 was restored by the addition of each of Ni $^{2+}$, Mn $^{2+}$, Ca $^{2+}$, Mg $^{2+}$, and Co $^{2+}$ up to 55%, 119%, 84%, 90%, and 87%, respectively (Supplementary Fig. 4a).

X-ray crystal structures of C-deglycosylation enzyme complexes. To elucidate the molecular basis of the enzymatic C-C bond cleavage reaction in the activation/inactivation of xenobiotic compounds by CGDs, we solved the apo structures of *PuCGD*, *EuCGD*, and AgCGD2, and the co-substrate structure of AgCGD2 with homoorientin, at 2.50, 2.40, 2.30 and 2.25 Å resolution, respectively (Fig. 3, Supplementary Fig. 9a and b and Supplementary Table 3). While these three CGDs shared similar heterodimer unit structures, *PuCGD* and *EuCGD* formed $\alpha_4\beta_4$ heterooctamers and AgCGD2 formed an $\alpha\beta$ heterodimer (Fig. 3, and Supplementary Fig. 9a and b). And Fig. 3a showed that homoorientin bound in a cavity at the interface between the α - and β -subunits of AgCGD2, indicating that the active site was created through heterodimerization.

Each β -subunit of both enzymes consisted of seven parallel β -sheets and formed a β -sandwich fold (Fig. 3a and Supplementary Fig. 9a and b). In contrast, the overall structure of *PuCGD $^{\alpha}$ (which represents the α -subunit of *PuCGD*) adopted a TIM-barrel fold, consisting of nine parallel β -strands assembled into a circular β -barrel, surrounded by a ring of solvent-exposed α -helices. A Dali search revealed that the *PuCGD $^{\alpha}$ structure exhibited moderate similarities with those of known sugar isomerases, including the xylose isomerase from *Planctopirus limnophila* (PDB ID: 4OVX) and the sugar-phosphate isomerase/epimerase from *Parabacteroides distasonis* ATCC 8503 (PDB ID: 3P6L), with Z-scores of 27.4 and 23.6, and RMSD values of 2.6 and 2.3 Å for the 270 and 262 Ca-atoms, respectively (19 and 17% amino acid sequence identities, respectively).**

The heterodimer structures of *EuCGD* and AgCGD2 were similar to that of *PuCGD*, with RMSD values of 2.2 and 3.5 Å, respectively (Fig. 3a–c, and Supplementary Fig. 9a and b). The interaction angle of AgCGD2 $^{\alpha}$ and AgCGD2 $^{\beta}$ was clearly different

from that of *PuCGD*, while the AgCGD2 $^{\alpha}$ structure was almost identical to that of *PuCGD $^{\alpha}$.*

The major structural differences among these enzymes lay on the interface between the α - and β -subunits; the α -subunits of *PuCGD* and AgCGD2 (represented as *PuCGD $^{\alpha}$ and AgCGD2 $^{\alpha}$, respectively) were composed of a TIM barrel domain and an additional domain with four α -helices (lid domain), which was not observed in *EuCGD $^{\alpha}$ (Fig. 3a and Supplementary Fig. 9a and b). Instead, the loop between M121–N133 of *EuCGD $^{\alpha}$, forming the active site cavity, was six residues longer than those of *PuCGD $^{\alpha}$ and AgCGD2 $^{\alpha}$. Moreover, the N-terminal loops of *EuCGD $^{\beta}$ and AgCGD2 $^{\beta}$ were also short as compared to that of *PuCGD $^{\beta}$. The loop of *PuCGD $^{\beta}$ formed the entrance of the cavity, while the corresponding loops of *EuCGD $^{\beta}$ and AgCGD2 $^{\beta}$ extended in different directions (Fig. 3a–c, and Supplementary Fig. 9a and b). On the other hand, 11 residues were inserted in the loop between $\beta 7$ and $\beta 8$ of *EuCGD $^{\beta}$ (T110–V139), as compared to those of *PuCGD $^{\beta}$ and AgCGD2 $^{\beta}$, and the loop was involved in the formation of the active site cavity.**********

Cryo-EM structures of C-deglycosylation enzyme complexes.

To obtain more structural details of C-deglycosylation enzymes, and to examine the flexibility of the lid-domain and the loop regions in *PuCGD* and *EuCGD*, we also determined the structures of *PuCGD* and *EuCGD* at 2.85 and 2.54 Å resolution, respectively, using single-particle cryo-EM (Supplementary Fig. 10 and Supplementary Table 4). The $\alpha_4\beta_4$ heterooctamer structures of both enzymes were almost identical to those of the X-ray crystal structures (RMSD values 0.6 and 0.4 Å, respectively) (Fig. 4). Intriguingly, however, the densities of the lid domain in *PuCGD $^{\alpha}$ and the long loops in *EuCGD* were disordered in the cryo-EM structures. These results suggested that open-to-closed conformational changes of the lid domain occurred upon substrate binding. Moreover, a comparison between the crystal structures of apo AgCGD2 and its homoorientin complex revealed that the lid domain of AgCGD2 moved ~ 3 Å and rotated ~ 18 degrees toward the active site upon substrate binding (Fig. 3b).*

Truncation of the lid domain of *PuCGD $^{\alpha}$ completely abolished the C-C bond cleavage activity, but did not affect protein expression or $\alpha_4\beta_4$ heterooctamer formation (Supplementary Fig. 3a and c). Although the *EuCGD* complex did not possess a lid domain, the disordered loop regions would play a similar role in the active site formation (Fig. 4b).*

Active site architecture. The crystal structure of AgCGD2 complexed with homoorientin (C6-glycosylated flavone) indicated that the active site was located between the α - and β -subunits.

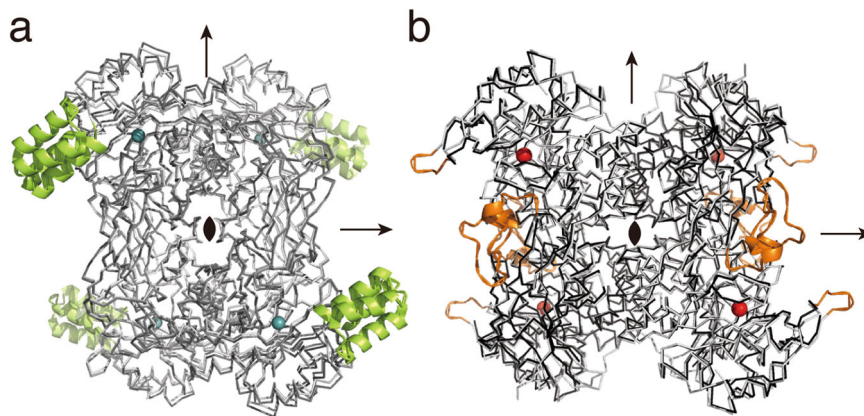


Fig. 4 Comparison of overall structures between X-ray crystal structures and cryo-EM structures. Comparison of X-ray crystal structures (pale gray) and cryo-EM structures (gray) of (a) *PuCGD* and (b) *EuCGD* (front view). Missing residues in cryo-EM structures are shown in (a) lime and (b) orange. Arrows represent approximate 2-fold axes.

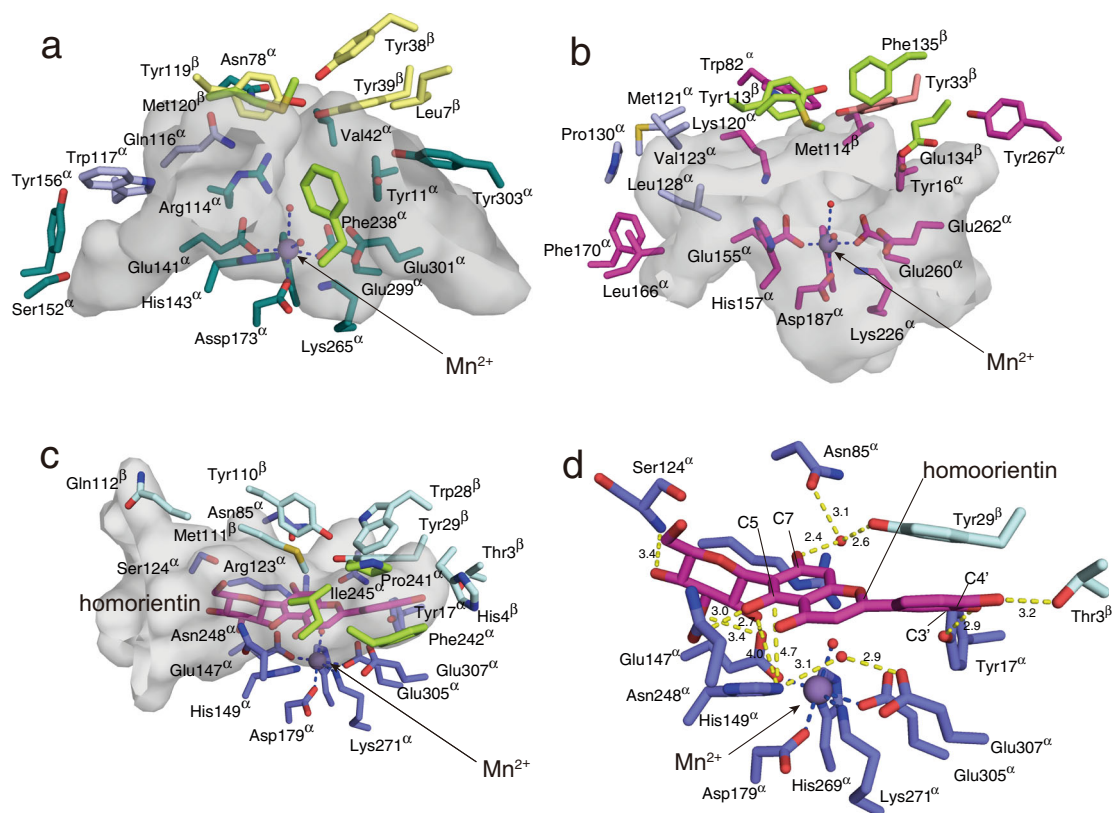


Fig. 5 Comparison of the active site architectures of C-deglycosylation enzymes. a–c Comparison of the active sites of (a) *PuCGD*, (b) *EuCGD*, and (c) *AgCGD2* with homoorientin (colored in magenta). Side views of the active site cavities are depicted as gray surfaces. d Interactions between homoorientin and active site residues in *AgCGD2*. The hydrogen bonds are shown by dashed yellow lines.

The difference anomalous Fourier maps of *AgCGD2*, *PuCGD*, and *EuCGD* suggested that these enzymes contain Mn^{2+} as a metal ion in the active site (Supplementary Fig. 11). The manganese ion was coordinated by Glu147 $^{\alpha}$ (which represents Glu147 in the α -subunit of *AgCGD2*), Asp179 $^{\alpha}$, His269 $^{\alpha}$, Lys271 $^{\alpha}$, Glu305 $^{\alpha}$, and a water molecule (Figs. 3c and 5). In the *PuCGD* structure, although a water molecule coordinated to the metal ion instead of Lys271 $^{\alpha}$, the positions of the bound metal and the residues 1His/2Glu/1Asp/1Lys were well conserved in the three homologs (Figs. 3c and 5, and Supplementary Fig. 9c and d). Notably, the mutations of these residues to alanine in *PuCGD* and *MiCGD* (E147 $^{\alpha}$ A, D179 $^{\alpha}$ A, H269 $^{\alpha}$ A, K271 $^{\alpha}$ A, and E305 $^{\alpha}$ A

mutants [the numbering of each of which follows the residues in *AgCGD2*]) dramatically reduced the C–C bond cleavage activity, indicating that metal-binding was essential for the enzyme reaction (Fig. 6). Although the CGDs accepted various metals to exhibit their activities, these observations suggested that the enzymes utilize Mn^{2+} as a native metal ion. On the other hand, the occupancy of the metal ion was relatively low. Considering the results of the metal dependency experiment showing that chelate reagent-treated enzymes exhibited higher activities on the addition of the metal ion than non-treated enzymes, some CGDs lacked metals in the heterologous expression or purification steps. This partial metal occupancy may affect enzyme activity as a

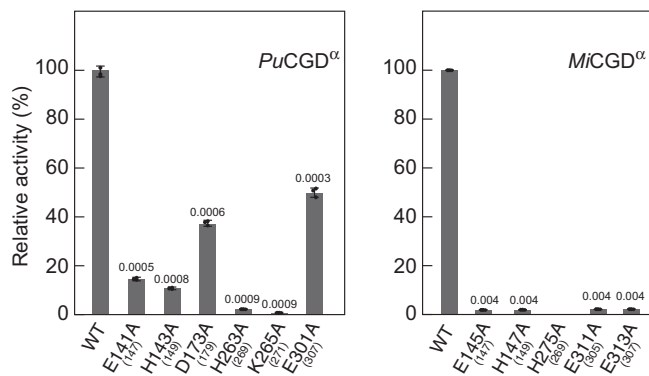


Fig. 6 Mutagenesis analysis of PuCGD α and MiCGD α . The numbers in brackets represent the corresponding numbers of the residues in AgCGD2. All experiments were repeated independently more than three times, and similar results were obtained. Statistical comparisons between means for the wild type (WT) and each mutant were performed by Student's t-test (2-tailed). The p-values were shown above each bar. The bars are means of $n = 3$ independent experiments (each data point was indicated by a black circle) and error bars indicate standard deviations. Data are presented as mean values \pm SD. All experiments were repeated independently three times with similar results.

result and cause small errors in the coordination sphere of the metal ions in the crystal structure.

While the metal-coordinating residues were only in the α -subunit, the aglycone moiety of homoorientin interacted with the residues in both subunits; for example, the hydroxyl groups at C3', C4', and C5 interacted with Thr3 β , Tyr17 α , and Asn248 α , respectively, and the hydroxyl groups at C2'' and C3'' of the sugar moiety interacted with Glu147 α , and the hydroxyl group at C4'' interacted with Ser124 α (Fig. 5d). On the other hand, the docking model of AgCGD2 with orientin (C8-glycosylated flavone), constructed based on the structure of a complex with homoorientin, suggested that the active site pocket of AgCGD2 is large enough to accommodate orientin with a similar sugar-binding mode (Supplementary Fig. 12a and b). In this model, the C7 and C3' of the aglycone moiety interacted with Arg123 α and the mainchain of Gln112 β , respectively.

The structure of AgCGD2 complexed with homoorientin indicated that the aglycon moiety was firmly fixed through a hydrogen-bonding interaction and surrounded by hydrophobic residues in the active site cavity, and that there were no space to change the conformation of the aglycon part for C6-glycosides; the conformational change would be necessary for this reaction to proceed, because the aglycon is dearomatized and the C6 or C8 carbon atom, which connects with glycoside, is changed to an sp³ from an sp² carbon in the enzyme reaction. In contrast, the docking model of AgCGD2 with orientin suggested that the enzyme possessed enough space for alteration of the conformation in the active site for C8-glycosides. These would be the possible reasons why AgCGD2 acted on C8-glycosylated compounds, rather than C6-glycosylated ones.

Sequence alignment among AgCGD2, MiCGD, AgCGD1, and MtCGD revealed that the residues that interact with hydroxyl groups of the aglycon of homoorientin in the AgCGD2 structure (including Tyr17 α and Asn248 α) are well conserved (Supplementary Fig. 13). On the other hand, the loop in AgCGD2 α comprising Pro241 α and Phe242 α is shortened, and Met249 α , which formed the active site cavity in AgCGD2, is substituted with a small amino acid, Ala or Gly, in the other enzymes. Moreover, the loop region between β 7 and β 8 (R107-G120 in AgCGD2 β), which is located close to the aglycon moiety, was not conserved and four residues were found to be inserted in

MiCGD β , AgCGD1 β , and MtCGD β (Supplementary Fig. 13). These differences would alter the shape of the active site and allow MiCGD, AgCGD1, and MtCGD to act on C6-glycosides rather than C8-glycosylated compounds.

We also constructed docking models of PuCGD and EuCGD with orientin and homoorientin, based on the co-crystal structure of AgCGD2 with homoorientin, to determine the structural basis of the substrate specificity of these enzymes (Supplementary Fig. 12c-f). The models suggested that PuCGD and EuCGD possess significant space to accept orientin and homoorientin, respectively (Supplementary Fig. 12d and e). On the contrary, PuCGD did not accept homoorientin because Tyr303 α and Leu7 β clash with the B-ring of homoorientin in the model (Supplementary Fig. 12c). In the model of EuCGD with orientin, furthermore, the B-ring of orientin also clashed with active site residues including Leu128 α and Pro115 β (Supplementary Fig. 12f). These observations are consistent with the substrate specificities of PuCGD and EuCGD (Figs. 1 and 2).

As described above, the active site was formed by the amino acid residues from each subunit: the α -subunit bound metal, which was essential for the reaction, and the β -subunit was involved in the substrate binding. To investigate whether the α - and β -subunits of each C-deglycosylation enzyme complex are exchangeable, we performed a swapping experiment on the β -subunits to construct PuCGD α -EuCGD β and EuCGD α -PuCGD β . The pull-down assays revealed that the β -subunits were not co-eluted with the α -subunits in both cases, indicating that the subunits of each C-deglycosidase were not exchangeable for the heterodimer formation (Supplementary Fig. 3d); in other words, the complex structure of each enzyme was crucial for its specific enzymatic activity.

Identification of catalytic residues. It has been proposed that 3''-oxo-puerarin is nonenzymatically isomerized to the 2''-oxo form in a basic buffer, to generate 2''-ene-3''-diol-puerarin as a putative intermediate^{16,17}. Considering the slightly acidic environment of a typical human gastrointestinal tract, and the optimal pH of the PuCGD and EuCGD enzyme reactions, however, one or more basic residues could facilitate the deprotonation from the C2'' position of the sugar moiety and the dearomatization of the A-ring in the flavonoid to catalyze the C-C bond cleavage reaction. Indeed, conserved His149 α (in AgCGD2) is situated near the homoorientin in AgCGD2, suggesting that His149 α can be involved in the catalytic reaction. The crystal structure of AgCGD2 in complex with homoorientin showed that distances of ND_{His149}-C2''_{homoorientin} and NE_{His149}-C6''_{homoorientin} were 4.2 Å and 4.7 Å, respectively (Supplementary Fig. 12a and b). Moreover, the conserved Glu307 α interacted with His149 α via a water molecule in the active site of AgCGD2. A similar hydrogen bond network was also observed in the active site of PuCGD. Notably, the docking model of AgCGD2 with orientin (C8-glycosylated flavone) suggested that the ND atom of His149 α was located at a distance of 3.8 Å from the C2'' atom of orientin, which is close enough to abstract the C2'' hydrogen atom. Furthermore, a water molecule coordinating to the metal ion was positioned close to the C7 hydroxyl group at a distance of 3.5 Å (Supplementary Fig. 12b).

To investigate the roles of these residues, therefore, the conserved His and Glu residues in the C8-glycoside-specific PuCGD and the C6-glycoside-specific MiCGD were substituted with Ala. The enzyme reaction with 3''-oxo-puerarin revealed that the activities of the H143 α A and E301 α A mutants of PuCGD (which corresponded to H149 and E307 in AgCGD2, respectively) were reduced to 10.4% and 49.8%, respectively (Fig. 6). Furthermore, the H147 α A and E313 α A mutants of MiCGD

showed dramatically reduced C-C bond cleavage activity (to 1.8 and 1.9%, respectively), as shown in Fig. 6. These findings suggested that the histidine residue facilitated the reaction as a base catalyst to deprotonate the C5-hydroxyl group of the aglycone and the C2 position of the sugar moiety via a water molecule. Meanwhile, the glutamic acid residue would support the positioning of the water molecule through hydrogen bond interactions.

Discussion

Microorganisms are involved in the degradation of natural and some artificial compounds, the resultant material resources being returned for use in nature²³. However, only limited numbers of metabolic pathways for various natural or artificial compounds have been identified. In previous studies, we identified the novel microbial metabolism of and metabolic enzymes for artificial compounds such as nitrile²⁴ and isonitrile²⁵, and natural compounds such as sesamin and curcumin^{26,27}.

Glycosides are ubiquitous compounds in nature. Plants synthesize glycosylated flavonoids, which are transferred to vacuoles or the cell wall²⁸, and microorganisms synthesize glycosylated antibiotics such as erythromycin²⁹, vancomycin³⁰, lankamycin³¹, lincosamycin³², and avermectin³³. We daily ingest plant-derived glycosylated compounds from vegetables and fruit. Those glycosides have been reported to be metabolized by gut microbes in the large intestine⁵; the resulting aglycones are absorbed from the intestine and show various bioactivities in our body³⁴. While a lot of microbes catalyzing C-glycoside deglycosylation reactions for various flavonoids and terpenoids have been reported^{5–11}, to the best of our knowledge, all of the so far known microbes were intestinal bacteria and the catalytic mechanisms for the reaction (involving the combination of active amino acid residues and a metal ion) remain unclear. In organic chemistry, on the other hand, C-C bond cleavage is a large issue because of its inert property. Although some strategies such as the use of transition-metals or generation of radicals have enabled to activate a C-C bond³⁵, the cleavage of the inherently inert bond is still challenging.

In this study, we characterized two C-glycoside deglycosylation enzymes (CGDs) from intestinal bacteria and three CGDs from soil bacteria, which were isolated through assimilation screening and genome mining. Further detailed biochemical analysis of CGDs from intestinal and soil bacteria revealed that these enzymes catalyzed selective C-C bond cleavage reactions toward 3'-oxo-C6- or C8-glycosylated compounds, while they exhibited broad substrate specificity toward the aglycone structures. Furthermore, SEC-MALS analysis revealed CGDs from intestinal bacteria were $\alpha_4\beta_4$ heterooctamer proteins, while those from soil bacteria were $\alpha\beta$ heterodimer proteins, which could be the reason why the optimal temperatures of intestinal bacteria (60 °C) were higher than those of soil bacteria (40 °C, Supplementary Fig. 4). These differences in subunit structures may be due to the difference in the bacterial phyla: *Microbacterium* and *Arthrobacter* belong to the phylum Actinomycetes, while intestinal *E. cellulosolvens* and the PUE strain belong to the phylum Firmicutes.

Through the crystallographic analysis of CGDs, we identified the relationships between the enzyme structure and function of CGDs. (i) The active site of CGD was formed by the interface between the α - and β -subunits and the complex structure of each enzyme was crucial for enzymatic activity. (ii) The conformational change of the lid domain would be important for the formation of the active site for catalysis. (iii) The metal ion was bound in the α -subunit, while several residues in the β -subunit were involved in the binding of a substrate. The effects of chelating agents and site-directed mutagenesis analysis revealed that

metal-binding was essential for the enzyme reaction. (iv) The shape of the active site was determined by the combination of α - and β -subunits, which plays a critical role in the substrate specificities of CGDs. Based on the crystal structures together with the results of biochemical and site-directed mutagenesis analyses, we propose a detailed mechanism for the C-C bond cleavage reactions of the unique C-deglycosylation enzyme complexes, as follows (Fig. 7 and Supplementary Fig. 14). (a) After the substrate binds to the active site with a conformational change of the lid domain, and the sugar moiety binds close to the coordinated metal and His149^a, a metal-hydroxide ion abstracts a proton from the hydroxyl group at the ortho or para position of the glycosylated carbon in the aglycone of the substrate. And then deprotonation of NE of His149^a occurs to dearomatize the aglycone moiety, followed by abstraction of a hydrogen atom from the C2'' position of the glycoside by His149^a as a base catalyst to form the 2''-ene-2'', 3''-diol intermediate. The hydrogen bond interactions among His149^a, Glu307^a, and a water molecule, and deprotonated NE of His149^a could facilitate the abstraction of the C2'' hydrogen atom. (b) The C-C bond between the aromatic ring and the sugar moiety is cleaved through a β -elimination-like reaction, to generate 1,5-anhydro-D-erythro-hex-1-en-3-ulose and the aglycone.

While C-C bond cleavage reactions have been reported for the biosynthesis or metabolism of natural products such as steroids, lignans, and fatty acids, these reactions are catalyzed by oxidative enzymes, including P450, radical SAM enzyme, and flavoproteins³⁶. In contrast, although the C-deglycosylation reaction is also initiated by oxidation of the C3-position of a glycoside by oxidoreductase, the C-C bond cleavage reaction of CGDs is a metal-assisted general acid/base mechanism. O-deglycosylation with a β -elimination-like reaction has been observed in the reactions of some glycoside hydrolase enzyme families³⁷, which catalyze the oxidation of the C3-hydroxyl groups of glycosides, followed by a nonenzymatic β -elimination-like reaction to cleave the C-O bonds in O-glycosides. In contrast, the cleavage of the inert C-C bond shown here requires *Pu*CGD homologs after C3-oxidation by oxidoreductase, indicating the importance of the dearomatization and subsequent deprotonation steps for the C-C bond cleavage by the unique C-deglycosylation enzyme complexes.

As described above, CGDs and microbes that perform bio-conversion of C-deglycoside compounds to aglycone have been identified in only intestinal bacteria¹⁸. However, together with our microbial screening, a database search suggested that CGD homologs are widely distributed in various bacterial phyla (Supplementary Fig. 1), indicating the universality of rare C-C bond cleavage reactions in nature. Considering that we identified *Mi*CGD from a C-glycoside-catabolizing microorganism on assimilation screening, the biological function of the C-deglycosylation reaction would be involved in the uptake of sugar as a carbon source from natural glycosylated compounds. Further identification of C-deglycosylation enzymes and investigation of the functions of other enzymes in the gene cluster will provide insights into the metabolic cycle of glycosylated compounds.

In conclusion, our structure-function analysis of C-deglycosylation enzymes revealed the overall structures, active site architecture, and key structural changes. Based on these structural observations and mutagenesis experiments, the C-C bond cleavage mechanism involving acid/base catalysis was proposed. We also indicated the generality of the reaction in both soil and intestinal microorganisms through biochemical and structural studies. Because the C-C bond cleavage is crucial for C-glycosides to exert their biological activities, these findings will facilitate clarification of the bio-availability of xenobiotic C-glycosides in humans and the biogeochemical circulation of C-glycosides in

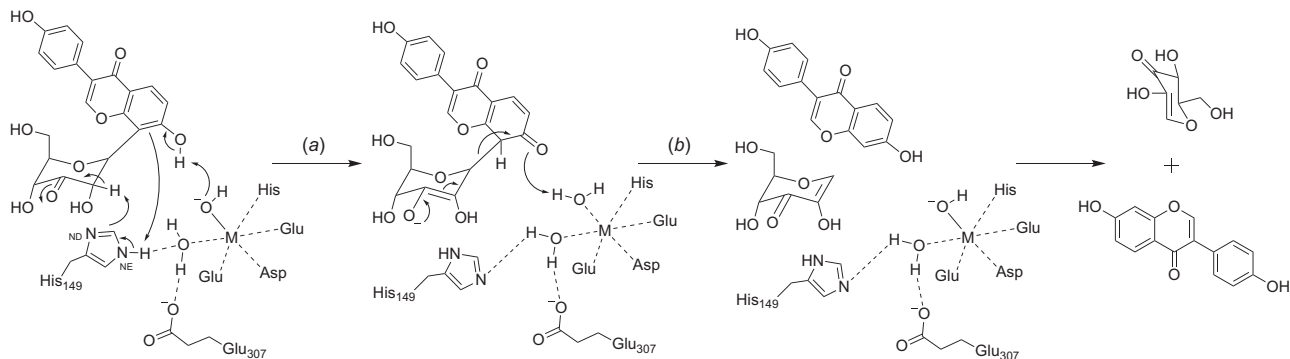


Fig. 7 Proposed mechanism of C-deglycosylation reaction. The numbering of amino acid residues in this figure is that for AgCGD2. M represents a metal ion.

nature. Future biochemical and structural analyses of the enzymes from intestinal bacteria will provide further insights into the molecular basis of the activation of prodrugs.

Methods

General remarks. Oligonucleotide primers were purchased from Eurofins Genomics. Other chemicals were purchased from Wako Chemical Ltd., and Kanto Chemical Co. Inc., (Tokyo, Japan). The LCMS data were obtained using a compact microTOF-MS (Bruker) attached to an LC-20AD UHPLC system (Shimadzu) with a COSMOSIL 2.5C18-MS-II column (2 mm i.d. \times 75 mm; Nacalai Tesque, Inc.). Analytical HPLC was performed on a Shimadzu LC20-AD HPLC system, using a Thermo Scientific Hypersil GOLD analytical column (4.6 \times 250 mm, 5 μ m). 3-Oxo-glucose and 3''-oxo-puerarin were synthesized according to the published method¹⁷.

Cloning and heterologous expression of DgpA, DfgE, PuCGD, EuCGD, MiCGD, AgCGD1, AgCGD2, and MtCGD. DNA fragments encoding DgpA, DgpB, DgpC, DfgA, DfgB, and DfgE were synthesized after codon optimization for expression in *E. coli* by FASMAC Co., Ltd (Supplementary Table 6). The full length of each gene was amplified using the corresponding primers listed in Supplementary Table 5. For DgpA, PuCGD, and EuCGD, each PCR product was cloned into the linearized pET22b vector (Merck Millipore) for expression as fusion proteins with a C-terminal His₆-tag. For the construction of the PuCGD and EuCGD plasmids, the amplified DgpB and DgpC fragments were tandemly ligated into the linearized pET22b vector. The amplified DfgE was inserted into the digested pET28a-MBP vector. For MiCGD, AgCGD1, AgCGD2, and MtCGD, each PCR product was cloned into the linearized pET24a(+) vector by the In-Fusion protocol (Clontech Laboratories, Inc.).

After sequence confirmation, the vectors containing the target genes were transformed into *E. coli* BL21(DE3) competent cells (in the case of DgpA, DfgE, PuCGD, and EuCGD) or *E. coli* Rosetta2 (DE3) cells (in the case of MiCGD, AgCGD1, AgCGD2, and MtCGD). The cells harboring the plasmids were cultured at 37 °C in LB medium, containing 100 μ g ml⁻¹ ampicillin (DgpA, DfgE, PuCGD, and EuCGD) or 2 \times YT media containing 50 μ g/ml kanamycin and 30 μ g/ml chloramphenicol (MiCGD, AgCGD1, AgCGD2, and MtCGD). To express the target proteins, isopropyl β -D-thiogalactopyranoside (IPTG) was then added to 0.2 mM (DgpA, DfgE, PuCGD, and EuCGD) or 0.5 mM (MiCGD, AgCGD1, AgCGD2, and MtCGD) (final concentration) when the OD₆₀₀ reached 0.6, and the cultures were continued for 20 h at 16 °C (DgpA, DfgE, PuCGD, and EuCGD) or 18 °C (MiCGD, AgCGD1, AgCGD2, and MtCGD). All of the following procedures were conducted at 4 °C.

For purification of DgpA, PuCGD, and EuCGD, the cultured cells were harvested by centrifugation at 6,000 \times g and resuspended in 50 mM HEPES (pH 7.6), containing 300 mM NaCl, 10% glycerol, and 10 mM imidazole (lysis buffer). For cell lysis, 1 mg ml⁻¹ lysozyme was added to the solution, followed by incubation at 4 °C for 1 h with slow rotation. The cells were then further disrupted by sonication on ice. The insoluble debris was removed by centrifugation at 12,000 \times g for 45 min. The supernatant was loaded onto a HisPur™ Ni-NTA Resin (Thermo Fisher Scientific) column, which was then washed with 50-column volumes (CVs) of lysis buffer containing 20 mM imidazole. Then, the enzymes were eluted with 3 CVs of lysis buffer containing 500 mM imidazole. The purified proteins were then applied to a Resource Q column and the bound proteins were eluted with a linear gradient of 0.05–1 M NaCl in 50 mM HEPES buffer (pH 7.6). The pooled protein from the Resource Q column was further purified by gel-filtration chromatography on a Superose 6 column (GE Healthcare), which was eluted with 20 mM HEPES (pH 7.6) buffer containing 100 mM NaCl. The concentration of each protein was calculated by measuring the ultraviolet absorption at A₂₈₀³⁸.

For the purification of MiCGD, AgCGD1, AgCGD2, and MtCGD, the pellet (20 g) was resuspended in twenty milliliters of 20 mM Tris-HCl buffer (pH 8.0). The cells were disrupted by sonication, as described above. The lysate was centrifuged at 27,000 \times g at 4 °C for 20 min. The recombinant proteins were purified using a His Trap HP column (GE Healthcare).

For the purification of DfgE, after His-tag purification, the purified fusion protein was dialyzed for 16 h in cold buffer containing 50 mM HEPES (pH 7.6), 300 mM NaCl, 10% glycerol, and 0.5 mM EDTA with the addition of TEV protease. The resulting protein solution was loaded twice onto a Ni-NTA resin column pre-equilibrated with 50 mM HEPES (pH 7.6), 300 mM NaCl, and the flow-through fraction was collected. The purified proteins were concentrated with Amicon centrifugal filter devices (30 KDa MWCO, Millipore), and used for *in vitro* assays.

General enzyme assays of wild type and mutants of PuCGD with synthesized 3''-oxo-puerarin. For enzyme assays, 0.1 mM 3''-oxo-puerarin was incubated in 50 mM KPi buffer (pH 7.4) containing 10 μ M purified recombinant PuCGD or mutants. The enzyme reactions were performed at 37 °C for 30 min and then quenched with 50 μ L of MeOH. The reaction mixtures were analyzed by HPLC. All measurements were conducted in triplicate.

General enzyme assays of PuCGD, EuCGD, MiCGD, AgCGD1, AgCGD2, and MtCGD with oxidoreductases. For enzyme assays of PuCGD and EuCGD, 0.1 mM substrate and 0.2 mM 3-oxo-glucose were incubated with 10 μ M DgpA or DfgE and PuCGD, EuCGD, or variants in 50 mM KPi buffer (pH 7.4). The enzyme reactions were performed at 37 °C for 30 min and then quenched with 50 μ L of MeOH. The reaction mixtures were analyzed by HPLC. All measurements were conducted in triplicate. For enzyme assaying of MiCGD, AgCGD1, AgCGD2, and MtCGD, the 100 μ L reaction mixture consisted of each of these enzymes, 10 mM Tris-HCl (pH 8.0), and 0.1 mM substrate, and was incubated at 28 °C. The reaction was stopped by adding an equal volume of acetonitrile. The amounts of reaction products were determined by HPLC-PDA or LC-MS.

HPLC and LC/MS analyses. A sample was applied to a Cosmosil π NAP column (4.6 \times 150 mm; Nacalai Tesque Co., Inc., Kyoto, Japan). HPLC and LC/MS analyses were performed using a Prominence system with a photodiode array detector (SPD-M20A) and an LCMS-8040 (Shimadzu, Kyoto, Japan). The HPLC conditions were as follows: flow rate, 1 ml min⁻¹; solvent A, 0.1% (v/v) HCOOH; and solvent B, methanol. After column equilibration with 50% solvent B, a linear gradient system of solvent B (50% to 100%) was applied over 13 min, followed by 100% solvent B for 2 min.

Temperature- and pH-optimization assays. To determine the optimal assay parameters for PuCGD, 0.1 mM 3''-oxo-puerarin was incubated in 50 mM KPi buffer (pH 7.4) containing 10 μ M purified recombinant PuCGD. For the EuCGD reaction, 0.1 mM homoorientin and 0.2 mM 3-oxo-glucose were incubated with 20 μ M DfgE in 50 mM KPi buffer (pH 7.4) for 2.5 h, and then 20 μ L of a 3''-oxo-homoorientin-containing solution was used as the substrate of EuCGD. The reactions were performed at 25, 30, 37, 42, 50, 60, and 70 °C for 1 h, and then quenched with 50 μ L of MeOH. The reaction mixtures were analyzed by HPLC. For thermal dependency estimation of MiCGD, AgCGD1, AgCGD2, and MtCGD, the 100 μ L reaction mixture consisted of 1 μ L of 1.1 mg ml⁻¹ MiCGD (or 1.9 mg ml⁻¹ AgCGD1, 0.38 mg ml⁻¹ AgCGD2, or 0.063 mg ml⁻¹ MtCGD), 1 μ L of 1 M Tris-HCl (pH 8.0), and 1 μ L of 10 mM carminic acid (or homoorientin for AgCGD1 and MtCGD, and orientin for AgCGD2). The experiments were performed in triplicate at 20, 25, 30, 35, 40, 45, 50, 60, and 70 °C for 60 min (or 15 min for AgCGD1 and MtCGD, and 5 min for AgCGD2). A Chill Heat CHT-101 (IWAKI Asahi Techno Glass, Tokyo, Japan) was used for the incubations from 10 to 50 °C, and a Dry

Thermo Unit DTU-1B (TAITEC, Tokyo, Japan) was used for the incubations at 60 and 70 °C. The amounts of reaction products were determined by HPLC-PDA.

To optimize the reaction pH of *PuCGD* and *EuCGD*, the enzymatic reactions were carried out in various reaction buffers with pH values ranging from 4.0–6.0 (citric acid-sodium citrate buffer), 6.0–8.5 (K_2HPO_4 - KH_2PO_4 buffer), 7.0–9.0 (Tris-HCl buffer), and 9.0–11.0 (CAPSO). All measurements were conducted in triplicate. For pH dependency estimation, the 100 μ l reaction mixture consisted of 1 μ l of 1.1 mg ml⁻¹ *MiCGD* (or 3.8 mg ml⁻¹ *AgCGD1*, 3.8 mg ml⁻¹ *AgCGD2*, or 3.2 mg ml⁻¹ *MtCGD*), 25 μ l of 0.4 M Britton-Robinson buffer [pH 6–8 (0.5 pH units)], and 2 μ l of 10 mM carminic acid (or homoorientin for *AgCGD1* and *MtCGD*, and orientin for *AgCGD2*). The experiments were performed for 60 min (or 120 min for *AgCGD1* and 15 min for *AgCGD2*) at 28 °C. The reaction was stopped by adding 100 μ l of acetonitrile. The amounts of reaction products were determined by HPLC-PDA.

Metal ion dependence experiment. The purified recombinant *PuCGD*, *MiCGD*, *AgCGD1*, and *AgCGD2* complex was treated with 5 mM EDTA and 5 mM Tiron at 4 °C with slow rotation for 3 days, and then the excess EDTA was removed by dialysis with 20 mM HEPES (pH 7.6) buffer containing 100 mM NaCl overnight. For *EuCGD*, the enzyme was incubated in buffer comprising 20 mM HEPES (pH 7.6) buffer, 100 mM NaCl, 20 mM EDTA, and 20 mM Tiron for 3 days at 4 °C. The enzyme reactions were performed without or with 1 mM divalent metal ion (Mn^{2+} , Ni^{2+} , Mg^{2+} , Fe^{2+} , Cu^{2+} , Co^{2+} , Zn^{2+} , and Ca^{2+}). All measurements were conducted in triplicate. The reactions were terminated with MeOH and the mixtures were centrifuged at 20,000 \times g for 10 min for further HPLC analysis.

Substrate specificity. Each of the following compounds was examined for substrate specificity analysis at a final concentration of 0.1 mM: carminic acid, mangiferin, homoorientin, isovitexin, puerarin, and orientin. Each of these compounds was added to the standard assay mixture. The production of the reaction products was detected by LC/MS. The reaction products of carminic acid were determined by mass spectrometry and NMR spectroscopy. Other reaction products were identified by mass spectrometry. For kinetic analysis of *PuCGD* and *EuCGD*, enzymatically prepared substrates (0.63, 1.3, 2.5, 5, 10, and 20 mM for 1, 3, and 4, and 0.16, 0.32, 0.63, 1.3, 2.5, and 5 mM for 2 and 5) were incubated with 2 μ M of each enzyme for 2 min at 30 °C (only for 2, 5 μ M of *PuCGD* was incubated for 90 min due to the low activity). And for kinetic analysis of *MiCGD*, *AgCGD1*, *AgCGD2*, and *MtCGD*, enzymatically prepared substrates (0.25, 0.5, 1, 2, 5, 10, and 20 mM) were incubated with 2 μ M of each enzyme for 30 min at 30 °C. Each reaction was performed in triplicate. GraphPad Prism 9 (GraphPad Software Inc., San Diego, CA) or Sigma Plot 12.5 (Systat Software Inc., San Jose, CA) was used for statistical data analysis.

Draft genome sequence of a carminic acid-metabolizing microorganism. Total DNA from *Microbacterium* sp. 5-2b, which was isolated by an enrichment method using carminic acid as a sole carbon source, was prepared as follows. The strain was cultured at 28 °C for 48 h in 100 ml of 1/10 \times 2YT media [0.1% (w/v) Bacto Yeast Extract (DIFCO), 0.16% (w/v) Bacto Tryptone (DIFCO), and 0.05% (w/v) NaCl]. Cells were harvested by centrifugation, washed with 10 mM Tris-HCl buffer (pH 8.0) containing 1 mM EDTA and 100 mM NaCl, and then suspended in 10 ml of 50 mM Tris-HCl buffer (pH 8.0) containing 10 mM EDTA and 15% (w/v) sucrose. The suspension was incubated with 7 mg ml⁻¹ of lysozyme at 37 °C for 3 h, and then 2 ml of 0.5 M EDTA (pH 8.0), 2 ml of 10% SDS and 2.7 mg of proteinase K were added to the solution, which was incubated at 55 °C for 16 h. DNA was purified by extracting the lysate with phenol/chloroform/isoamyl alcohol (25/24/1; v/v/v), followed by precipitation with isopropanol, treatment with RNase, and then reprecipitation with ethanol. Draft genome sequencing of strain 5-2b was performed using an Illumina HiSeq platform (Hokkaido System Science Co., Ltd., Sapporo, Japan). We obtained 44.6 million reads of a 100 bp paired-end read. A total of 477 contigs comprising 189 ~1,692,717 bp were assembled. The draft genome sequence was annotated with MiGAP (<http://www.migap.org>).

MiCGD homologs. Using the protein sequence of *MiCGD*^a (Supplementary Table 6) as a query, a BLAST search was performed against the database of the nucleotide collection. Among the protein sequences of the top 100 ORFs from the BLAST results, we chose ten microorganisms with genome data available online. Among them, the two *MiCGD*^a homologs from *Arthrobacter globiformis* NBRC12137, and one *MiCGD*^a homolog from *Microbacterium trichothecenolyticum* NBRC15077 were designated as *AgCGD1*, *AgCGD2*, and *MtCGD*, respectively. They were cloned and heterologously expressed in *E. coli* BL21 Star (DE3).

Size-exclusion chromatography-multiangle static light scattering. SEC-MALS analyses were performed with a WTC-030S5 (WYATT Technology) using a LaChrom Elite high-performance liquid chromatography system (Hitachi). Light scattering and the refraction index were measured using a Dawn Heleos II detector (Wyatt Technology) and a RI-101 detector (Shodex), respectively. The column was equilibrated at 20 °C with 20 mM Tris-HCl buffer, pH 8.0, containing 100 mM NaCl. Samples (2 mg ml⁻¹) were injected at the buffer flow rate of 0.5 mL/min. The

obtained data were recorded and processed using the ASTRA 6.1 software (Wyatt Technologies).

Crystallization and structure determination. *PuCGD* crystals were obtained at 20 °C in 24% (w/v) PEG4000, 100 mM Tris-HCl (pH 8.5) with 15 mg ml⁻¹ of purified *PuCGD*, by the sitting-drop vapor-diffusion method. *EuCGD* crystals were obtained at 20 °C in 1190 mM $(NH_4)_2SO_4$, 100 mM Tris-HCl (pH 8.5) with 15 mg ml⁻¹ of purified *PuCGD*, by the sitting-drop vapor-diffusion method. The crystallization conditions for *AgCGD2* were initially screened using Crystal Screen 1 and 2 (Hampton Research), Wizard Screens I and II (Rigaku), PEGsII (Qiagen), Index (Hampton Research), PEGIon/PEGIon2 (Hampton Research), and a Protein complex suite (Qiagen) with a Protein Crystallization System 2 (PXS2) at the Structural Biology Research Center, High Energy Accelerator Research Organization, Japan³⁹. Diffraction quality crystals were obtained in 25–30% (w/v) PEG4000, 0.1 M MES (pH 6.5), and 0.2 M potassium iodide at 20 °C. The crystals of *PuCGD* and *EuCGD* were transferred into the cryoprotectant solution (reservoir solution containing 25% (v/v) glycerol). The crystals of *AgCGD2* were cryoprotected by immersion in a solution containing 15% (w/v) PEG 1000, 21% (w/v) PEG4000, 60 mM MES (pH 6.5), and 60 mM potassium iodide.

The X-ray diffraction data sets of *PuCGD* and *EuCGD* were collected at -178 °C using a beam wavelength of 1.08 Å at BL-1A (Photon Factory, Tsukuba, Japan). The diffraction data sets were processed and scaled using the XDS program package and Aimless of ccp4^{40,41}. X-ray diffraction data of *AgCGD2* for the native SAD method were collected at -178 °C, using an Eiger X16M detector on BL-17A of the Photon Factory, KEK (Tsukuba, Japan). The diffraction data were processed and scaled by XDS and XSCALE, respectively⁴¹. The phases were determined using the Crank2 program, by the native SAD method⁴². Crystallographic refinement and model building were performed using PHENIX.refine⁴³ and Coot⁴⁴. During the crystallographic refinement, a strong peak was found for the active site. A difference anomalous Fourier map that was calculated using diffraction data collected at (1.8900 Å) wavelength of manganese showed a strong peak at the active site. However, no peaks were found in the difference anomalous Fourier map calculated from the low remote (1.9000 Å) wavelength data.

Crystals of the homoorientin complex were prepared by soaking crystals of *AgCGD2* in a crystallization solution containing 10 mM homoorientin, for 2 h. Crystals of the homoorientin complex were cryoprotected with a solution containing 10 mM homoorientin, 15% (w/v) PEG 1000, 21% (w/v) PEG4000, 60 mM MES (pH 6.5), and 60 mM potassium iodide, for 15 s. Diffraction data sets of *AgCGD2* were collected at -178 °C using a Pilatus 2M-F detector on NE3A of the PF-AR, KEK (Tsukuba, Japan). Data processing and scaling were performed as described above.

The initial phase of the *PuCGD* structure was determined by molecular replacement, using the cryo-EM structure of *PuCGD* as the search model. Molecular replacement was conducted with Phaser in PHENIX⁴⁵. The initial phase was further calculated with AutoBuild in PHENIX⁴⁶. The crystal structure of the *AgCGD2*-homoorientin complex was determined by the molecular replacement method, using Molrep⁴⁷. Crystallographic refinement was performed using PHENIX.refine⁴³ and Coot⁴⁴. The final crystal data and intensity statistics are summarized in Supplementary Table 3. The Ramachandran statistics are as follows: 96.5% favored, 3.5% allowed for *PuCGD* and 96.8% favored, 3.2% allowed for *EuCGD*, 97.3% favored, 2.7% allowed for *AgCGD2* with homoorientin. All crystallographic figures were prepared with PyMOL (DeLano Scientific, <http://www.pymol.org>).

The three-dimensional model structure of orientin was generated with the CHEM3D ULTRA 10 program (Cambridge Soft), and the geometries were optimized with the elbow tool in phenix⁴⁸. Orientin or homoorientin was manually added to the active site to fit the aglycone part of homoorientin, with Coot⁴⁴. The parameters of orientin for the refinement were obtained with the PRODRG server (<http://davapc1.bioch.dundee.ac.uk/prodrng/>).

Cryo-EM sample preparation and data collection. The purification of *PuCGD* and *EuCGD* for cryo-EM analysis was performed in the same manner as for crystallization. The concentrations of *PuCGD* and *EuCGD* were adjusted to 37.2 μ M and 44.8 μ M (concentration of heterooctomer). For cryo-grid preparation, 3 μ l samples were applied onto a holey carbon grid (Quantifoil, Cu, R1.2/1.3, 300 mesh), which was rendered hydrophilic by a 30 sec glow-discharge in air (11 mA current) with a PIB-10 plasma ion bombarder (Vacuum Device Inc., Ibaraki, Japan). The grid was blotted for 5 sec at force 20 for *PuCGD* and 20 sec at force 0 for *EuCGD* in 100% humidity at 18 °C, and then plunged into liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific). The cryo-EM data collection was conducted using a Talos Arctica microscope (Thermo Fisher Scientific) at 200 kV in the nanoprobe mode, using the EPU software at the cryo-EM facility in KEK (Ibaraki, Japan). For *PuCGD*, the movies were collected by a 4k \times 4k Falcon 3EC direct electron detector (electron counting mode) at a nominal magnification of 120,000 \times (0.88 Å/pixel), with an accumulated dose of 50 electrons per Å² over 49 frames. For *EuCGD*, the movies were collected by the Falcon 3EC direct electron detector at a nominal magnification of 150,000 \times (0.69 Å/pixel) magnification, with an accumulated dose of 50 electrons per Å² over 62 frames.

Cryo-EM data processing software. For the cryo-EM analysis, motion correction and dose-weighting were performed using the MotionCor2 frame alignment program⁴⁹. The contrast transfer function (CTF) parameters were estimated using Gctf⁵⁰ and CTFIND4⁵¹. The particles were picked using SPHIRE-crYOLO with the general model⁵². The ctflimit function⁵³ implemented in a Python module, morphology.py, of SPARX/SPHIRE^{54,55} was used to calculate the smallest box size that ensures no CTF aliasing in the reciprocal space, up to the target resolution for a given defocus value. RELION3⁵⁶ and RELION 3.1.0-beta were used for all other SPA steps: reference-free 2D classification, ab initio reconstruction, 3D classification, 3D refinement, CTF refinement for the refinement of higher-order aberrations, anisotropic magnification, per-particle defocus and beam tilt, and Bayesian polishing for beam-induced motion corrections⁵⁷. After each of the 3D refinements, the “gold-standard” FSC resolution with the 0.143 criterion⁵⁸ in RELION was used as a global resolution estimation with phase randomization, to account for possible artifactual resolution enhancement caused by solvent mask^{59,60}. The local resolutions of the 3D cryo-EM maps were estimated using RELION’s own implementation. UCSF Chimera⁶¹ and e2display.py of EMAN2⁶² were used for the visualization of the output 2D/3D images. The detailed methods for cryo-EM processing are written in *Supplementary Methods*.

Mutagenesis of PuCGD and MiCGD. The plasmids expressing the mutant enzymes of PuCGD and MiCGD were constructed with a QuikChange Site-Directed Mutagenesis Kit (Stratagene) and a KOD-plus mutagenesis kit (Toyobo Co., Ltd., Osaka, Japan), respectively, according to the manufacturer’s protocol. The variants of PuCGD and MiCGD were transformed into *E. coli* BL21(DE3) and BL21-Star (DE3), respectively. The expression, purification, and enzyme reactions of all variants were performed in the same manners as for the wild-type enzymes. The primers used for the construction of mutants are listed in Supplementary Table 5.

Preparation of manual docking model structures. The three-dimensional model structures of orientin and homoorientin were generated by the CHEM3D ULTRA 10 program (Cambridge Soft), and their geometries were optimized with the elbow tool in phenix. Orientin or homoorientin was manually added in the active site to fit the position of the C-C bond between C-glycoside and aglycone. Then, the conformation of ligands was manually modified to avoid the close contacts between the ligand and the active site residues. Here, the positional relationship among the C-C bond, His149, and Glu307 (numbering of AgCGD2), was defined to be almost the same in all models. It is noted that we did not modify the conformation of the active site residues.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Source data are provided with this paper. The data generated in this study are provided in the Supplementary Information/Source Data file. The crystallographic data that support the findings of this study are available from the Protein Data Bank (<http://www.rcsb.org>). The coordinates and the structure factor amplitudes for the apo structures of PuCGD, EuCGD, AgCGD2, and AgCGD2 complexed with homoorientin were deposited under accession code 7EXZ⁶³, 7EXB⁶⁴, 7DNN⁶⁶, and 7DNN⁶⁶, respectively. The cryo-EM maps and the atomic coordinates for PuCGD and EuCGD determined by cryo-EM have been deposited in Electron Microscopy Data Bank (EMDB, <https://www.ebi.ac.uk/pdbe/emdb/>) and PDB with accession codes EMD-30808 and EMD-30809, and 7DRD⁶⁷ and 7DRE⁶⁸, respectively. Source data are provided with this paper.

Received: 30 March 2021; Accepted: 12 October 2021;

Published online: 02 November 2021

References

- Thursby, E. & Juge, N. Introduction to the human gut microbiota. *Biochem. J.* **474**, 1823–1836 (2017).
- Kobashi, K., Akao, T., Hattori, M. & Namba, T. Metabolism of drugs by intestinal bacteria. *Bifidobact. Microflora* **11**, 9–23 (1992).
- Koppel, N., Rekdal, V. M. & Balskus, E. P. Chemical transformation of xenobiotics by the human gut microbiota. *Science* **356**, 1246–1257 (2017).
- Wei, B. et al. Discovery and mechanism of intestinal bacteria in enzymatic cleavage of C-C glycosidic bonds. *Appl. Microbiol. Biotechnol.* **104**, 1883–1890 (2020).
- Hur, H. G., Lay, J. O., Beger, R. D., Freeman, J. P. & Rafii, F. Isolation of human intestinal bacteria metabolizing the natural isoflavone glycosides daidzin and genistin. *Arch. Microbiol.* **174**, 422–428 (2000).
- Braune, A., Engst, W. & Blaut, M. Identification and functional expression of genes encoding flavonoid O- and C-glycosidases in intestinal bacteria. *Environ. Microbiol.* **18**, 2117–2129 (2016).
- Seyed Hameed, A. S., Rawat, P. S., Meng, X. & Liu, W. Biotransformation of dietary phytoestrogens by gut microbes: A review on bidirectional interaction between phytoestrogen metabolism and gut microbiota. *Biotechnol. Adv.* **43**, 107576 (2020).
- Braune, A. & Blaut, M. Deglycosylation of puerarin and other aromatic C-glycosides by a newly isolated human intestinal bacterium. *Environ. Microbiol.* **13**, 482–494 (2011).
- Sanugul, K. et al. Isolation of a human intestinal bacterium that transforms mangiferin to norathyriol and inducibility of the enzyme that cleaves a C-glycosyl bond. *Biol. Pharm. Bull.* **28**, 1672–1678 (2005).
- Braune, A. & Blaut, M. Intestinal bacterium *Eubacterium cellulosolvens* deglycosylates flavonoid C- and O-glycosides. *Appl. Environ. Microbiol.* **78**, 8151–8153 (2012).
- Nakamura, K. et al. The C-glycosyl bond of puerarin was cleaved hydrolytically by a human intestinal bacterium strain PUE to yield its aglycone daidzein and an intact glucose. *Chem. Pharm. Bull.* **59**, 23–27 (2011).
- Robert, J. & Rivory, L. Pharmacology of irinotecan. *Drugs Today* **34**, 777–803 (1998).
- Takasuna, K. et al. Involvement of β -glucuronidase in intestinal microflora in the intestinal toxicity of the antitumor camptothecin derivative irinotecan hydrochloride (CPT-11) in rats. *Cancer Res* **56**, 3752–3757 (1996).
- Billign, T., Griffith, B. R. & Thorson, J. S. Structure, activity, synthesis and biosynthesis of aryl-C-glycosides. *Nat. Prod. Rep.* **22**, 742–760 (2005).
- Kytidou, K., Artola, M., Overkleeft, H. S. & Aerts, J. M. F. G. Plant glycosides and glycosidases: a treasure-trove for therapeutics. *Front. Plant Sci.* **11**, 357 (2020).
- Rauter, A. P., Lopes, R. G. & Martins, A. C-Glycosylflavonoids: identification, bioactivity and synthesis. *Nat. Prod. Commun.* **2**, 1175–1196 (2007).
- Nakamura, K., Zhu, S., Komatsu, K., Hattori, M. & Iwashima, M. Expression and characterization of the human intestinal bacterial enzyme which cleaves the C-glycosidic bond in 3'-oxo-puerarin. *Biol. Pharm. Bull.* **42**, 417–423 (2019).
- Nakamura, K., Zhu, S., Komatsu, K., Hattori, M. & Iwashima, M. Deglycosylation of the isoflavone C-glycoside puerarin by a combination of two recombinant bacterial enzymes and 3-oxo-glucose. *Appl. Environ. Microbiol.* **86**, e00607–e00620 (2020).
- Zhou, Y., Zhang, H. & Peng, C. Puerarin: A review of pharmacological effects. *Phytother. Res.* **28**, 961–975 (2014).
- Hlima, B. H., Bejar, S., Riguet, J., Haser, R. & Aghajari, N. Identification of critical residues for the activity and thermostability of *Streptomyces* sp. SK glucose isomerase. *Appl. Microbiol. Biotechnol.* **97**, 9715–9726 (2013).
- Fujiwara, T., Saburi, W., Matsui, H., Mori, H. & Yao, M. Structural insights into the epimerization of β -1,4-linked oligosaccharides catalyzed by cellobiose 2-epimerase, the sole enzyme epimerizing non-anomeric hydroxyl groups of unmodified sugars. *J. Biol. Chem.* **289**, 3405–3415 (2014).
- Bosshart, A., Hee, C. S., Bechtold, M., Schirmer, T. & Panke, S. Directed divergent evolution of a thermostable D-tagatose epimerase towards improved activity for two hexose substrates. *ChemBioChem* **16**, 592–601 (2015).
- Jenkins, A. H., Schyns, G., Potot, S., Sun, G. & Begley, T. P. A new thiamin salvage pathway. *Nat. Chem. Biol.* **3**, 492–497 (2007).
- Kobayashi, M. & Shimizu, S. Metalloenzyme nitrile hydratase: Structure, regulation, and application to biotechnology. *Nat. Biotechnol.* **16**, 733–736 (1998).
- Sato, H., Hashimoto, Y., Fukatsu, H. & Kobayashi, M. Novel isonitrile hydratase involved in isonitrile metabolism. *J. Biol. Chem.* **285**, 34793–34802 (2010).
- Kumano, T., Fujiki, E., Hashimoto, Y. & Kobayashi, M. Discovery of a sesamin-metabolizing microorganism and a new enzyme. *Proc. Natl Acad. Sci. USA* **113**, 9087–9092 (2016).
- Hassaniasab, A., Hashimoto, Y., Tomita-Yokotani, K. & Kobayashi, M. Discovery of the curcumin metabolic pathway involving a unique enzyme in an intestinal microorganism. *Proc. Natl Acad. Sci. USA* **108**, 6615–6620 (2011).
- Le Roy, J., Huss, B., Creach, A., Hawkins, S. & Neutelings, G. Glycosylation is a major regulator of phenylpropanoid availability and biological activity in plants. *Front. Plant Sci.* **7**, 735 (2016).
- Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**, 176–178 (1990).
- Hubbard, B. K. & Walsh, C. T. Vancomycin assembly: nature’s way. *Angew. Chem. Int. Ed.* **42**, 730–765 (2003).
- Egan, R. S. & Martin, J. R. Structure of lankamycin. *J. Am. Chem. Soc.* **92**, 4129–4130 (1970).

32. Zhao, Q., Wang, M., Xu, D., Zhang, Q. & Liu, W. Metabolic coupling of two small-molecule thiols programs the biosynthesis of lincomycin. *Nature* **518**, 115–119 (2015).
33. Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. & Omura, S. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proc. Natl Acad. Sci. USA* **96**, 9509–9514 (1999).
34. Williamson, G. & Clifford, M. N. Role of the small intestine, colon and microbiota in determining the metabolic fate of polyphenols. *Biochem. Pharmacol.* **139**, 24–39 (2017).
35. Wang, B., Perea, M. A. & Sarpong, R. Transition metal-mediated C-C single bond cleavage: making the cut in total synthesis. *Angew. Chem. Int. Ed.* **59**, 18898–18919 (2020).
36. Kim, E. M., Seo, J. H., Baek, K. & Kim, B. G. Characterization of two-step deglycosylation via oxidation by glycoside oxidoreductase and defining their subfamily. *Sci. Rep.* **5**, 10877 (2015).
37. Guengerich, F. P. & Yoshimoto, F. K. Formation and cleavage of C-C bonds by enzymatic oxidation-reduction reactions. *Chem. Rev.* **118**, 6573–6655 (2018).
38. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423 (1995).
39. Kato, R., Hiraki, M., Yamada, Y., Tanabe, M. & Senda, T. A fully automated crystallization apparatus for small protein quantities. *Acta Crystallogr. Sect. F. Struct. Biol. Commun.* **77**, 29–36 (2021).
40. Kabsch, W. XDS. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 125–132 (2010).
41. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D.* **69**, 1204–1214 (2013).
42. Pannu, N. S. et al. Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr. Sect. D.* **67**, 331–337 (2011).
43. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D.* **68**, 352–367 (2012).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
45. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
46. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 213–221 (2010).
47. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
48. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. Sect. D.* **65**, 1074–1080 (2009).
49. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
50. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
51. Rohou, A. & Grigorieff, N. CTFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
52. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
53. Penczek, P. A. et al. CTER-Rapid estimation of CTF parameters with error assessment. *Ultramicroscopy* **140**, 9–19 (2014).
54. Hohn, M. et al. SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).
55. Moriya, T. et al. High-resolution single particle analysis from electron cryo-microscopy images using SPHIRE. *J. Vis. Exp.* 55448 (2017).
56. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, 1–22 (2018).
57. Zivanov, J., Nakane, T. & Scheres, S. H. W. A bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCr* **6**, 5–17 (2019).
58. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
59. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
60. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
61. Pettersen, E. F. et al. UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
62. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
63. DgpB-DgpC complex apo 2.5 angstrom. <https://doi.org/10.2210/pdb7EXZ/pdb>, (2021).
64. DfgA-DfgB complex apo 2.4 angstrom. <https://doi.org/10.2210/pdb7EXB/pdb>, (2021).
65. Crystal structure of the AgCarB2-C2 complex. <https://doi.org/10.2210/pdb7DNN/pdb>, (2021).
66. Crystal structure of the AgCarB2-C2 complex with homoorientin. <https://doi.org/10.2210/pdb7DNN/pdb>, (2021).
67. Cryo-EM structure of DgpB-C at 2.85 angstrom resolution. <https://doi.org/10.2210/pdb7DRD/pdb>, (2021).
68. Cryo-EM structure of DfgA-B at 2.54 angstrom resolution <https://doi.org/10.2210/pdb7DRE/pdb>, (2021).

Acknowledgements

This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (JSPS KAKENHI Grant Numbers JP16H06443, JP19K15703, JP19K05784, JP19H05687, JP20H00490, JP20K22700, and JP21K18246), the New Energy and Industrial Technology Development Organization (NEDO, Grant Number JPNP20011), the PRESTO program of the Japan Science and Technology Agency, and the Fuji Foundation for Protein Research. This work was also supported in part by Platform Project for Supporting Drug Discovery and Life Science Research (Basis of Supporting Innovative Drug Discovery and Life Science Research) from AMED (JP21am0101071 (support number 1553) to T.S., KEK).

Author contributions

T. Mori, T.K., I.A. and M. Kobayashi designed the experiments. T. Mori, T.K., H.H., S.W., S.H. and Y.T. performed the in vitro analysis and crystallization experiments. M.S. and T. Mori performed the X-ray crystallography. T. Mori, T. Moriya, N.A. and M. Kawasaki performed the Cryo-EM analysis. T. Mori, T.K., H.H., S.M., T. Moriya, Y.H., I.A. and M. Kobayashi analyzed the data. T. Mori, T.K., T.S., I.A. and M. Kobayashi wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-26585-1>.

Correspondence and requests for materials should be addressed to Toshiya Senda, Ikuro Abe or Michihiko Kobayashi.

Peer review information *Nature Communications* thanks Steven Hardwick, Ki Hyun Nam and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021