RSC Advances

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *RSC Adv.*, 2025, **15**, 28592

# Mechanistic elucidation of enzymatic *C*-glycosylation: facilitation by proton transfer to UDP-glucose

Daisuke Terada, [ab] Taichi Inagaki [a] and Miho Hatanaka *[ac]

*C*-Glycosyltransferases have garnered attention owing to their ability to synthesize *C*-glycosides with high conversion and selectivity in one-pot reactions. Their potential in rational enzyme engineering makes them valuable for the synthesis of diverse *C*-glycosides. However, the detailed reaction mechanism remains unclear. To address this, we investigated the *C*-glycosylation of phloretin catalyzed by the glycosyltransferase GgCGT in the presence of the coenzyme UDP-glucose. Using density functional theory (DFT) calculations on a cluster model, we identified the most favorable pathway for *C*-glycosylation. The reaction proceeds *via* an initial proton transfer from phloretin to UDP-glucose, followed by the nucleophilic attack of phloretin on the glucose moiety and subsequent dissociation of UDP in an $S_N2$-like manner. The $S_N2$ step yields a non-aromatic intermediate, which can be rapidly converted to *C*-glycoside even without an enzymatic environment. The key residue that facilitates the rate-determining $S_N2$ step is His-27, which stabilizes phloretin *via* hydrogen bonding. Additionally, to clarify why alternative products such as *O*-glycosides are not formed, we also investigated the *O*-glycosylation pathway. Our calculations revealed that *O*-glycosylation was promoted by proton transfer from UDP-glucose, like *C*-glycosylation, but was suppressed by structural fixation due to hydrogen bonding among phloretin, glucose, and GgCGT.

Received 15th April 2025
Accepted 6th August 2025

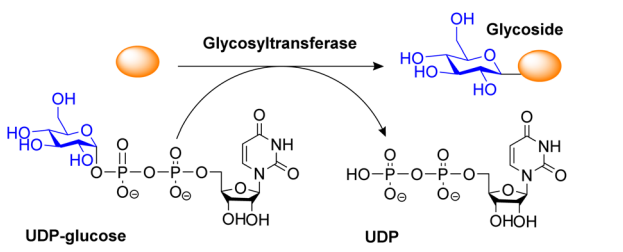
DOI: 10.1039/d5ra02643a

rsc.li/rsc-advances

Open Access Article. Published on 12 August 2025. Downloaded on 1/21/2026 5:27:40 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

## 1. Introduction

Glycosyltransferases are enzymes that utilize sugar nucleosides as coenzymes to transfer sugar moieties to acceptor molecules.[1] For instance, a glycosyltransferase that employs uridine diphosphate glucose (UDP-glucose) as the sugar nucleoside catalyzes the transfer of the glucose moiety to an acceptor molecule, resulting in the formation of a glycoside. Simultaneously, UDP-glucose is converted into UDP with the subsequent release of glycosides and UDP (Scheme 1). Glycosides have diverse bioactive properties and are widely used in the development of pharmaceuticals and cosmetics. They are classified as *O*-, *N*-, *S*-, or *C*-glycosides based on the atom to which the sugar moiety is attached. Among these, *C*-glycosides have garnered interest owing to their remarkable metabolic stability.[2] While numerous natural products containing *O*- and *N*-glycosides and their enzymatic reactions have been studied, *C*-glycosylated natural products remain relatively rare.[3,4] According to Putkaradze *et al.*, only 55 enzymes capable of producing *C*-glycosides have been identified, of which 30 have been discovered since 2020.[4] Consequently, the number of reported crystal structures of *C*-glycosyltransferases is limited, and comprehensive studies on their catalytic mechanisms are currently lacking. Glycosyltransferase-mediated *C*-glycosylation offers efficient one-pot syntheses of target *C*-glycosides with precise stereoselectivity.[5] Conversely, the organic synthesis of *C*-glycosides typically involves multi-step processes, including protection and deprotection steps, owing to the presence of hydroxyl groups in the sugar moiety.[6] Additionally, because glycosides can have up- and down-stereoisomers at the anomeric position, organic synthesis also requires a scheme to control stereoselective glycosylation. Therefore, controlling




Scheme 1 Enzymatic glycosylation with the coenzyme of UDP-glucose.

[a]*Graduate School of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa 223-8521, Japan. E-mail: miho_hatanaka@keio.jp*

[b]*Analysis Technology Center, FUJIFILM Corporation, 210, Nakanuma, Minamiashigara-shi, Kanagawa, 250-0193, Japan*

[c]*Institute for Molecular Science, 38 NishigoNaka, Myodaiji, Okazaki, Aichi 444-8585, Japan*



© 2025 The Author(s). Published by the Royal Society of Chemistry

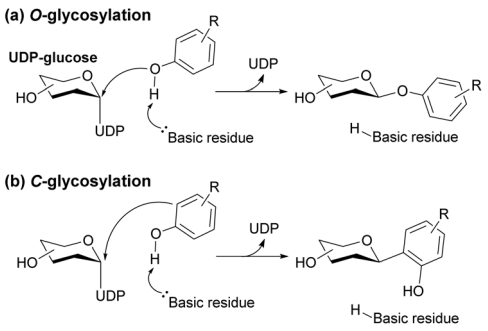
Open Access Article. Published on 12 August 2025. Downloaded on 1/21/2026 5:27:40 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

enzymatic reactions for the synthesis of *C*-glycosides is considered a more efficient alternative to conventional organic synthesis methods.[7] Understanding the mechanism of enzymatic *C*-glycosylation is crucial for rational enzyme engineering. Such advances enable the production of a broader range of *C*-glycosides, thereby expanding their potential applications.[8]

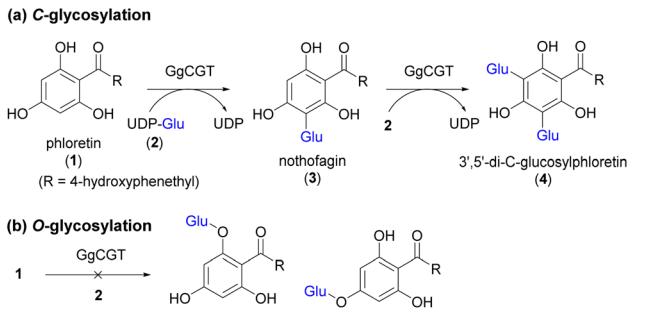
The reaction mechanisms underlying the glycosylation have predominantly been investigated for *O*-glycosyltransferases using quantum mechanical (QM) approaches.[9] These investigations demonstrated that basic amino acid residues within the active site of glycosyltransferases facilitate the activation of the hydroxyl group of the substrate. This activation subsequently promotes nucleophilic attack by the oxygen atom at the anomeric position, culminating in the formation of *O*-glycosidic bonds (Scheme 2a).[9,10] Although the precise reaction mechanism of *C*-glycosylation enzymes remains elusive, experimental evidence has provided hypotheses regarding the mechanism of *C*-glycosylation. Previously, a mechanism involving the rearrangement of *O*-glycosides to produce *C*-glycosides was proposed.[10] However, studies by Gatmann and Nidetzky on a *C*-glycosyltransferase, named OsCGT, demonstrated that *C*-glycosides are produced directly.[11] Furthermore, their research revealed that the activation of the hydroxyl group by a basic residue (His) in OsCGT is critical for *C*-glycosylation (Scheme 2b), and this activation process is analogous to that observed in the mechanism of *O*-glycosylation shown in Scheme 2a. Studies on another *C*-glycosyltransferase, PlCGT, also suggested that an Asn–Asp dyad functions as the basic active site residue.[12] It is postulated that activation of the hydroxyl group by this basic residue facilitates nucleophilic attack from an adjacent carbon atom, resulting in *C*-glycosylation (Scheme 2b).[4,11–13] Additionally, investigations into PlCGT revealed that mutations in residues surrounding the active site pocket, rather than the active residue itself, altered the *C*/*O*-selectivity of glycosylation.[12] This finding indicates that residues near the pocket also play a role in *C*-glycosylation. These studies highlighted a shared feature between *O*-glycosylation and *C*-glycosylation, namely the involvement of basic active site residues. However, the specific mechanism that governs selectivity between these glycosylation types remains unclear.

Among the *C*-glycosyl transferases, GgCGT, discovered in the medicinal plant Glycyrrhiza glabra, has been studied for its crystal structure and enzymatic activity.[14] GgCGT catalyzes *C*-glycosylation of phloretin (**1**), a polyphenol, under enough amount of UDP-glucose (**2**). As shown in Scheme 3, this reaction involves C–C bond formation between **1** and the glucose moiety in **2**, along with the dissociation of UDP, resulting in the production of mono-*C*-glycoside nothofagin (**3**). GgCGT also exhibits *C*-glycosylation activity for **3**, yielding the di-*C*-glycosylate, 3′,5′-*C*-glucosylphloretin (**4**). A time-course measurement of this enzymatic reaction revealed that **1** was initially converted to **3**, followed by the production of **4** over time, ultimately leading to the complete conversion of **1** into **4**.[14] The crystal structures of GgCGT in complex with different ligands, such as **2**, UDP with **1**, and UDP with **3** also supported stepwise di-*C*-glycosylation.[14] For instance, the position of **2** was fixed by seven H-bonds from GgCGT, and its anomeric carbon was found to be close to the reactive carbon (3′-C). The positions **1** and **3** in the complexes were similar, suggesting that the first and second *C*-glycosylation steps occurred similarly. These results indicated that glycosylation started with the coordination of **1** and **2** to the active pocket of GgCGT, followed by the first *C*-glycosylation, release of **3** and UDP from GgCGT, re-coordination of **3** and **2**, and a second *C*-glycosylation. Other glycosylated products, such as *O*-glycosylated phloretins, were not detected in the GgCGT-catalyzed reaction. Complete selectivity of di-*C*-glycosylation by GgCGT was also observed for phloretin derivatives containing a flopropione unit, which was considered to be a key factor controlling selectivity.[14]

In this study, we used density functional theory (DFT) calculations to gain insight into the reaction mechanism of GgCGT-catalyzed *C*-glycosylation and to determine the key to controlling selectivity between *C*- and *O*-glycosylation. We focused on the monoglycosylation step because di-glycosylation occurs in a stepwise manner, as mentioned above. We constructed a model complex involving the active site of GgCGT, a model substrate, and model UDP-glucose. We examined the reaction energy profile of GgCGT-catalyzed mono-*C*-glycosylation using a model complex and discussed how GgCGT promotes *C*-glycosylation. To gain deeper insight into the role of the basic residue His-27, we compared the energy profiles of mono-*C*-glycosylation catalyzed by GgCGT and its mutant, in which His-27 was replaced by alanine. Finally, we investigated the reaction profiles of the two types of mono-*O*-glycosylation




Scheme 2 Schematic illustration of the proposed mechanism of *O*-glycosylation[9,10] (a) and *C*-glycosylation[4,11–13] (b).




Scheme 3 (a) GgCGT-catalyzed *C*-glycosylation of phloretin (**1**) with UDP-glucose (**2**), affording mono- and di-*C*-glycosylates, **3** and **4**, respectively. (b) *O*-Glycosylation of **1** was not detected.

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 12 August 2025. Downloaded on 1/21/2026 5:27:40 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

and explained the key to controlling the selectivity for *C*- and *O*-glycosylation.

## 2. Computational methods

We constructed a model complex of GgCGT, phloretin **1**, and UDP-glucose **2** based on two crystal structures: one for GgCGT with few defects (PDB:6L5R) and another for GgCGT in complex with **2** (PDB:6L5P). The phloretin **1** was placed in the cavity near the glucose site. The protonation states of ionizable amino acid residues were determined using PROPKA v. 3.4.0 program[15] at pH 8.0, considering the experimental conditions of the GgCGT enzymatic reaction.[14] The protonation state of His-27 was manually determined to form an H-bond network with Asp-122, as shown in several glycosyltransferases.[14,16] Before extracting the model complex, we performed energy minimization of the entire system to avoid clashes, followed by MD simulations at 300 K for 100 ps in the NVT ensemble and for 100 ps in the NPT ensemble using GROMACS 5.1.5 program.[17] Corresponding parameters of ligands in the force field were generated using the general AMBER force field (GAFF2),[18] coupled with the Amber FF14SB force field[19] employed for GgCGT and the TIP3P model used for water molecules.[20] The restrained electrostatic potential (RESP)[21] charge was applied to establish the partial atomic charge of ligands, with the HF/6-31G(d) calculation through the Gaussian 16 package.[22] The geometry at the final snapshot was used to extract the model complex. To consider the enzymatic reaction field, we selectively included the following amino acid residues in the model complex: His-27, Asp-122, Thr-145, Ser-284, His-366, Asn-370, Asp-390, and Gln-391, which neighbored the binding sites of **1** and **2**. The structures of **1** and **2** were simplified to **1a** and **2a** by substituting *R* and *R*′ groups with methyl groups, as shown in Fig. 1. The positions of the main-chain atoms of GgCGT were fixed throughout the geometry optimization, except for the side-chain atoms His-27, Asp-122, Ser-284, His-366, and Asp-390, to avoid unrealistic conformational changes (see Fig. 2). This model has a total of 186 atoms, 92 of which are fixed during the geometry optimization. The total charge of the model is −4. To understand the effect of His-27, we compared the energy profile of the model cluster with that of the H27A mutant, in which His-27 of the wild type was replaced by Ala. The geometry of the H27A was generated by replacing the imidazole moiety of His-27 in the wild-type cluster model with an H atom, while keeping the positions of the fixed atoms unchanged. This modelling strategy assumes that a mutation reducing side-chain volume (H27A) is unlikely to induce large-scale structural rearrangements of the active site, as the overall protein fold is constrained by the backbone. All geometry optimizations and frequency calculations were performed at the B3LYP-D3/6-31G(d,p) level of theory, and the electronic energies were refined with the single-point calculations at the B3LYP-D3/6-311+G(2d,2p) using the PCM solvation model with the value of dielectric constant $\varepsilon = 4$, which was often used for representing enzymatic environment.[23] Note that we have confirmed that the choice of the basis set and the dielectric constant have little effect on the results (see Fig. S1 and Table S1, respectively). The Gibbs free energies were evaluated using the electronic energies by the single-point calculations and the Gibbs free energy correction terms at 298.15 K and 1 atm obtained by the frequency calculations. Both the Gibbs free energy difference $\Delta G$ and the electronic energy difference $\Delta E$ are shown in our study. All the obtained transition states (TSs) were confirmed by the intrinsic reaction coordinate (IRC) calculations.[24] All the geometry optimization, reaction path search, and IRC calculations for the model system were performed with the GRRM program[25] using energies and energy derivatives computed with the Gaussian 16 program.

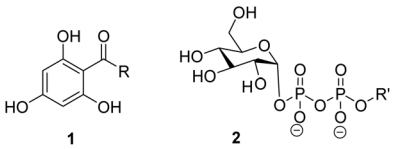



Fig. 1 Simplified **1** and **2**, named **1a** and **2a**, in which *R* and *R*′ groups were replaced with methyl groups.

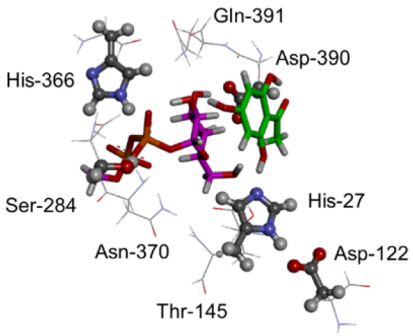



Fig. 2 Model complex of GgCGT/**1a**/**2a**. Atoms in ball-and-stick were optimized by fixing the atoms in line for all the geometry optimizations.

## 3. Results and discussion

### 3.1. Reaction profile of *C*-glycosylation

Fig. 3 shows the Gibbs free energy profile of mono-*C*-glycosylation of the model complex, as shown in Fig. 2. Initially, the model GgCGT formed a complex with **1a** and **2a**, designated **INT0**, through several H-bonds (indicated by the green dotted lines in Fig. 3a). The basic residue His-27 interacted with the phenolic hydroxyl group of **1a** at the ortho position (2′-OH; see **INT0** in Fig. 3a) and played a key role in maintaining the orientation of **1a**. The position of **2a** was stabilized by other residues, specifically Ser-284, His-366, and Asn-370, which anchored the phosphate group, and Thr-145 and Asp-390, which stabilized the glucose moiety through H-bonds, as observed in the crystal structure (PDB:6L5P). The fixed positioning of **1a** and **2a** established an H-bond network involving the 4′-OH of **1a**, 2-OH of glucose in **2a**, and phosphate of **2a** (see Fig. 4a). Proton transfer along this network (indicated by red arrows in $\mathbf{TS_{0-1}}$ in Fig. 3) took place with an activation barrier of 9.3 kcal mol$^{-1}$, leading to the formation of an unstable


© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 12 August 2025. Downloaded on 1/21/2026 5:27:40 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

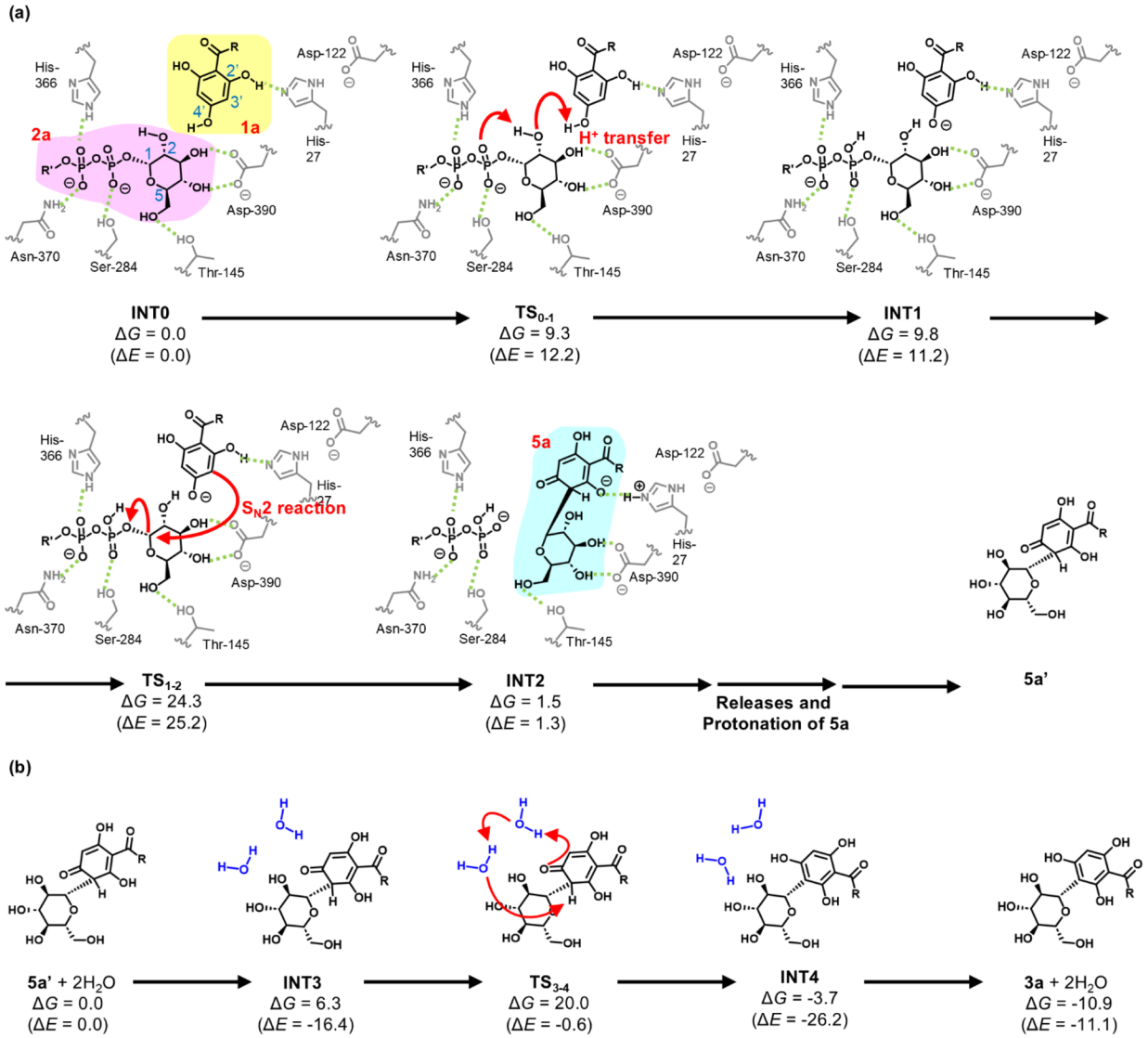



Fig. 3 Reaction profile of GgCGT catalyzed mono-*C*-glycosylation of the model protein **1a** with the model UDP-glucose **2a**, starting from proton transfer and $S_N2$ reaction (a), followed by the re-aromatization after the release of an intermediate **5a′** from GgCGT (b). Numbers in blue (in **INT0**) represent the position numbers. The Gibbs free energy and the electronic energies (in parentheses) are in kcal $mol^{-1}$. The energy references for (a) and (b) are their starting structures (*i.e.*, **INT0** and **5a′** + $2H_2O$, respectively). The single point calculations at B3LYP-D3/6-311+G (2d, 2p) with the PCM model were carried out at the geometries optimized at B3LYP-D3/6-31G(d,p).

intermediate, **INT1**. This proton relay occurred only between **1a** and **2a**, without the involvement of surrounding residues (Fig. 4a and b). This proton transfer from **1a** enhanced the nucleophilicity of the 3′-C of **1a**, which was oriented towards the anomeric C (1 C) of glucose, itself constrained by the surrounding residues (see Fig. 4b). Consequently, the nucleophilic attack of the 3′-C of **1a** to anomeric *C* of glucose could proceed, which induced the dissociation of the UDP moiety from glucose. Focusing on the transition state of this step (**$TS_{1-2}$**) in Fig. 4c, the distances between the reactive atoms (forming C–C bonds and cleaving C–O bonds) were equidistant, indicating an $S_N2$-type mechanism. The IRC calculation for **$TS_{1-2}$** indicated that UDP dissociation preceded C–C bond formation. This sequence suggests that proton abstraction from **1a** to the UDP moiety enhances the electrophilicity of UDP, leading to its dissociation. Therefore, we can say that the initial proton transfer (**INT0** → **INT1**) set the stage for the subsequent $S_N2$ reaction by increasing the nucleophilicity of **1a** and the electrophilicity of UDP. The $S_N2$ reaction through **$TS_{1-2}$** yielded a non-aromatic intermediate, **INT2**, which was 8.3 kcal $mol^{-1}$ more stable than **INT1**. In **INT2**, His-27 abstracts a proton from **1a**, which stabilizes the resulting cyclohexadienone moiety of **5a** (see Fig. 4d). Interestingly, proton transfer from **1a** to His-27 was not observed in **$TS_{1-2}$**, which implied that the anionic nature of the phloretin moiety in **INT1** hindered proton abstraction by His-27. Thus, the previous assumption about basic residue-mediated substrate activation (see Scheme 2b) is called into question. Further analysis of the role of His-27 is described in the following section.

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 12 August 2025. Downloaded on 1/21/2026 5:27:40 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

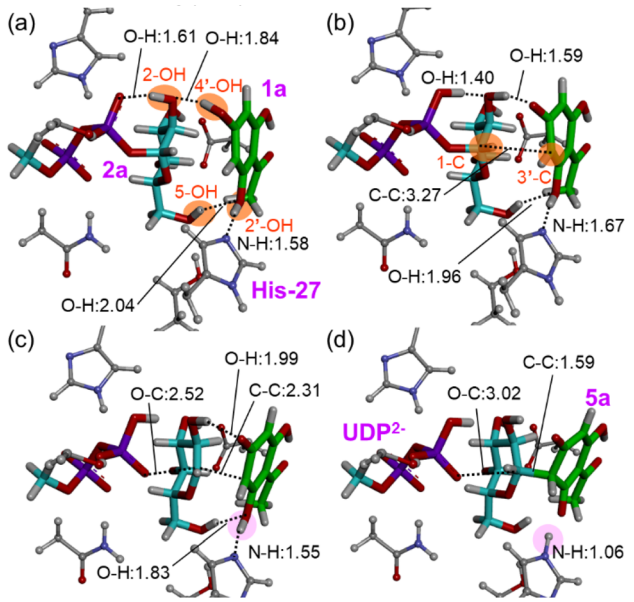



Fig. 4 Geometries around the reaction center in INT0 (a), INT1 (b), $TS_{1-2}$ (c), and INT2 (d). Numbers in black and orange represent atomic distances in Å and position numbers shown in INT0 in Fig. 3a, respectively. Atoms fixed during the geometry optimization were omitted. Their full views are also shown in Fig. S2.

After the formation of **5a**, the release of **5a** from GgCGT and the protonation of **5a** affording neutral **5a′** could take place. To compute such an energy profile, the effect of conformational changes in the entire enzyme and solvent molecules must be considered.[26] Though our model complex was insufficient to discuss the energy profile involving the release of **5a**, it has been experimentally shown that mono-*C*-glycosides are released from GgCGT and then re-entered for di-*C*-glycosylation as shown in Scheme 3a, which means that substrate could enter and leave GgCGT easily. Once the non-aromatic intermediate **5a′** was formed, the re-aromatization of **5a′** to yield mono-*C*-glycoside **3a** occurred *via* a proton relay with additional water molecules (see Fig. 3b). The activation barrier of this step, including two water molecules ($\mathbf{TS_{3-4}}$), was only 20.0 kcal mol$^{-1}$, indicating that this step could proceed even without an enzymatic environment.

In summary, the mono-*C*-glycosylation of **1a** proceeded in a stepwise manner, starting with proton transfer from **1a** to **2a**, followed by the $S_N2$ reaction involving C–C bond formation between **1a** and glucose, coupled with the dissociation of UDP, and then re-aromatization *via* proton transfer. The rate-determining step in the overall reaction is the $S_N2$ reaction, with an activation barrier of 24.3 kcal mol$^{-1}$. However, the fact that the mono-*C*-glycosylation completes within 10 minutes at 37 °C (ref. 14) allows for an experimental upper limit of the activation barrier to be estimated at approximately 20 kcal mol$^{-1}$. Thus, the activation barrier obtained from our calculations appears to be somewhat overestimated. This may be attributed to the artificial suppression of structural changes in the active site from the fixation of atoms in the model system. To consider this effect, we compared the **INT0** geometry of the cluster model with that of the entire enzyme optimized at the ONIOM(B3LYP-D3/Amber) level. As shown in Fig. S3, the geometries of the reaction center in the cluster model and the entire enzyme were similar; however, the geometries of the entire active site were slightly different due to the influence of fixed amino acid residues. Therefore, while these geometric constraints render our model less suitable for predicting quantitative activation energies, they do not invalidate the qualitative description of the reaction pathway. The primary finding of this study—that the reaction is initiated by a proton transfer from the substrate—is governed by the electronic interactions within the reaction center, which are well-described by our model. We thus conclude that our model is sufficient for elucidating the fundamental reaction coordinate. In addition, our proposed mechanism, in which proton transfer from phloretin to UDP-glucose promoted *C*-glycosylation, was appropriate from a chemical perspective, as it enhanced the nucleophilicity of phloretin and the leaving ability of UDP.

Next, to confirm that the reaction from **INT0** to **INT2** proceeds *via* a concerted mechanism, as shown in Fig. 3a, rather than a stepwise mechanism, we calculated the potential energy surface (PES) along the reaction coordinates (Fig. 5). The reaction coordinates for proton transfer (from **INT0** to **INT1**), and

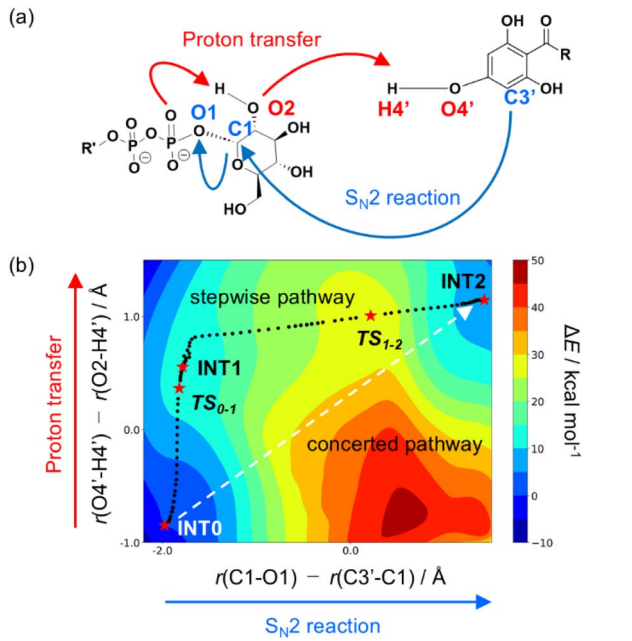



Fig. 5 Reaction coordinates from INT0 to INT2 and the atoms involved in the reaction (a). The potential energy surface (PES) $\Delta E$ (in kcal mol$^{-1}$) along with the reaction coordinates representing proton transfer and $S_N2$ processes in vertical and horizontal axes, respectively (b). The reaction coordinates of proton transfer and $S_N2$ reaction are defined as the distance differences $r$(O4′–H4′)–$r$(O2–H4′) and $r$(C1–O1)–$r$(C3′–C1), respectively. The PES was plotted using the single-point energies calculated at geometries optimized with fixed reaction coordinates, sampled at intervals of 0.1 Å (vertical axis) and 0.3 Å (horizontal axis). The single-point calculations were performed at the B3LYP-D3/6-311+G(2d,2p) with the PCM model. The geometry optimizations with constrained reaction coordinates were carried out at the B3LYP-D3/6-31G(d,p) level. The red stars represent the structures shown in Fig. 3a. The black dots indicate the IRC pathways from $TS_{0-1}$ and $TS_{1-2}$. The white line represents the concerted pathway.


© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 12 August 2025. Downloaded on 1/21/2026 5:27:40 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

the $S_N2$ reaction (from **INT1** to **INT2**) are represented by the differences in the bond distances formed and cleaved during the transformation from **INT0** to **INT2** (Fig. 5a). In Fig. 5b, the vertical axis represents the reaction coordinates of proton transfer, defined as the difference between the O4′–H4′ and H4′–O2 distances, whereas the horizontal axis represents the reaction coordinates for the $S_N2$ reaction, defined as the difference between the C3′–C1 and C1–O1 distances. The reaction pathway along the vertical axis followed by that along the horizontal axis describes a stepwise mechanism, whereas the pathway following the diagonal across the PES represents a concerted mechanism. The PES map reveals a significant energy barrier in the central region, effectively ruling out the possibility of a concerted mechanism. Additionally, a substantial barrier along the horizontal axis from **INT0** indicates that *C*-glycosylation does not begin with the $S_N2$ reaction. These findings confirmed that the identified pathway, in which proton transfer preceded the $S_N2$ reaction, accurately described the mechanism of *C*-glycosylation.

### 3.2. Effect of the basic residue His-27

To examine the effect of the basic residue His-27 on this reaction, we focused on the mutant H27A, in which His-27 in GgCGT was replaced with alanine (Ala-27). This H27A mutant was experimentally synthesized and was known to hardly catalyze the *C*-glycosylation of phloretin **1**.[14] We created a model of the H27A mutation (Fig. 6a) and calculated the energy profile from **1a** to **5a** through the proton transfer and $S_N2$ reaction. As shown in Fig. 6b, the formation of **5a** from **1a** was much more endothermic ($\Delta G = 8.9$ kcal mol$^{-1}$) with the H27A mutant than with the wild type (GgCGT). This indicates that proton abstraction by His-27, which did not occur in the H27A mutant, contributed to stabilizing **5a**. Additionally, the activation barrier at the rate-determining step ($S_N2$ reaction) with the H27A mutant was 3.1 kcal mol$^{-1}$ higher than that with the wild-type. To understand the origin of the difference in the activation barrier, energy decomposition analysis (EDA) was performed, which divided the activation barrier into contributions from interaction and deformation (see Fig. S4). The difference in the activation barrier was attributed to the interaction energy between the enzyme and the other compounds (**1a** and **2a**). This result indicated that His-27 acted as an activator as a basic residue, as well as a base to abstract a proton, leading to the formation of **5a**.

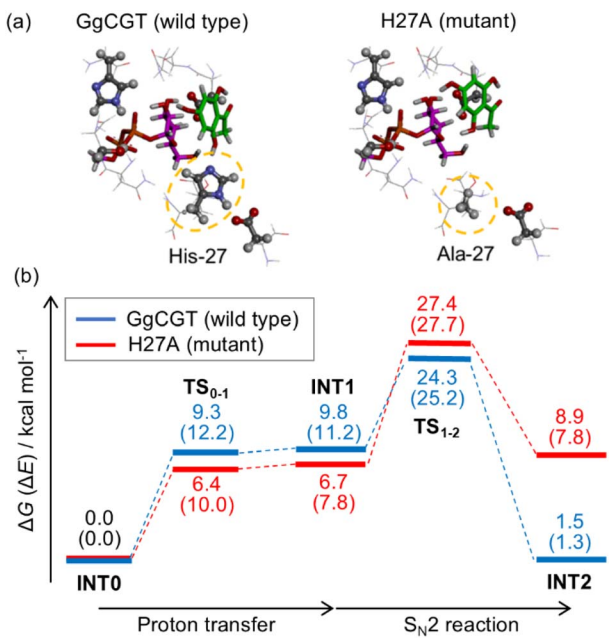



Fig. 6 The model GgCGT (wild type) and its mutant H27A, where His-27 in the GgCGT was replaced by alanine (a). The Gibbs free energy profile (in kcal mol$^{-1}$) of *C*-glycosylation of **1a** catalyzed by the model GgCGT and H27A (b). The electronic energies are in parenthesis. Energy references of the GgCGT and H27A systems are their starting structures, which correspond to **INT0** in Fig. 3a. The single point calculations at B3LYP-D3/6-311+G (2d, 2p) with the PCM model were performed at the geometries optimized at B3LYP-D3/6-31G (d, p).

### 3.3. Origin of the selectivity for *C*- and *O*-glycosylation

To understand the origin of the regioselectivity (*i.e.*, selectivity for *C*-glycosylation and *O*-glycosylation) for phloretin **1** in GgCGT, we investigated the reaction profile of *O*-glycosylation of the model phloretin **1a**, which does not actually proceed with GgCGT. The *O*-glycosylation could occur *via* two pathways: the nucleophilic attack on the anomeric C atom of glucose by the 2′-O and the 4′-O of **1a** (see Fig. 7a and b, respectively). The nucleophilic attack by 2′-O could be activated by His-27, as described in Scheme 2a. Though **INT0** exhibited an H-bond between 2′-OH and His-27 (see N–H bond distance of 1.58 Å in Fig. 4a), the activation barrier of the *O*-glycosylation was as high as 35.5 kcal mol$^{-1}$ ($\mathbf{TS_{0-5}}$ in Fig. 7c). The IRC calculation from $\mathbf{TS_{0-5}}$ revealed that the reaction *via* $\mathbf{TS_{0-5}}$ was a concerted pathway and started with proton transfer from **1a** to His-27, followed by the nucleophilic attack on anomeric C atom by 2′-O and then UDP dissociation (Fig. S5). The high activation barrier could be attributed to the lack of an H-bond network between **1a** and **2a**, which stabilizes **INT0**. Additionally, the negative charge was more localized on the UDP in $\mathbf{TS_{0-5}}$ and **INT5** than in $\mathbf{TS_{1-2}}$ and **INT2** because of the lack of proton transfer from **1a** to UDP (Fig. 3a), which could also cause the instability of $\mathbf{TS_{0-5}}$ and **INT5**.

Another possible *O*-glycosylation pathway, the nucleophilic attack by 4′-O of **1a**, could be activated by the proton abstraction from 4′-OH, *i.e.,* the formation of **INT1**. Thus, we attempted to obtain the TS of this nucleophilic attack while maintaining the H-bonds in **INT1**. However, such a structure was impossible owing to the strong fixation of relative orientation between **1a** and glucose *via* two H-bonds (see Fig. 4b for OH bonds in **INT1** with distances of 1.59 and 1.96 Å). To form the bond between 4′-O of **1a** and anomeric C of glucose, two H-bonds between **1a** and **2a** had to be dissociated (see OH bonds with 2.58 and 2.37 Å in Fig. 8a), resulting in an activation barrier of 25.6 kcal mol$^{-1}$ ($\mathbf{TS_{1-6}}$), which was 1.3 kcal mol$^{-1}$ higher than that of *C*-glycosylation through $\mathbf{TS_{1-2}}$ (Fig. 7c). The corresponding product **INT6**, which yielded neutral *O*-glycoside and $UDP^{2-}$, was 10.9 kcal mol$^{-1}$ less stable than **INT2** despite retaining the

Open Access Article. Published on 12 August 2025. Downloaded on 1/21/2026 5:27:40 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

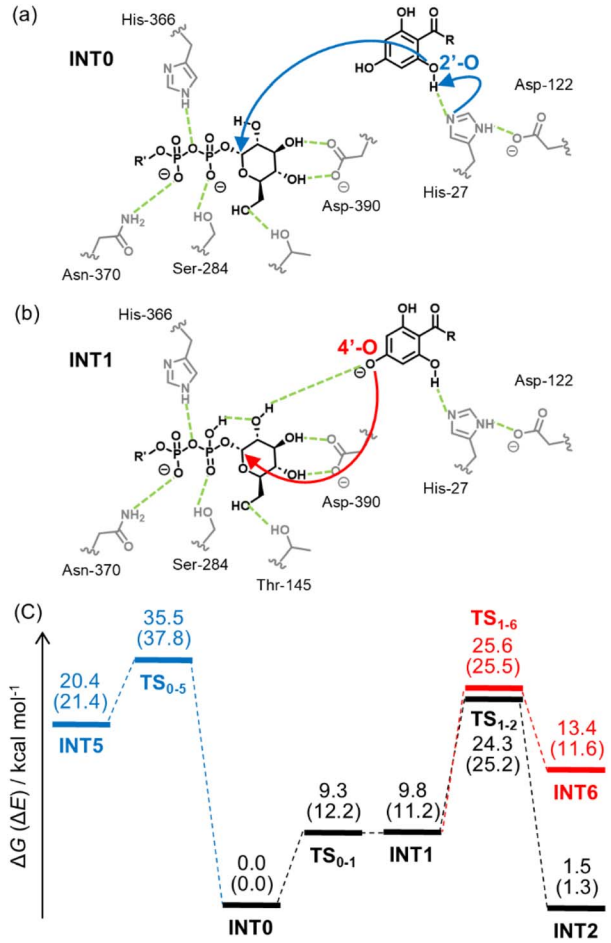



Fig. 7 Two possible GgCGT catalyzed *O*-glycosylation pathways: 4′-*O*-glycosylation where **INT1** formation precedes (a) and 2′-*O*-glycosylation starting from **INT0** (b). The Gibbs free energy profiles (in kcal mol$^{-1}$) of the 4′-*O*- and 2′-*O*-glycosylations are shown in red and blue, respectively (c). The electronic energies (in kcal mol$^{-1}$) are in parenthesis. The single point calculations at B3LYP-D3/6-311+G(2d,2p) with the PCM model were performed at the geometries optimized at B3LYP-D3/6-31G(d,p).

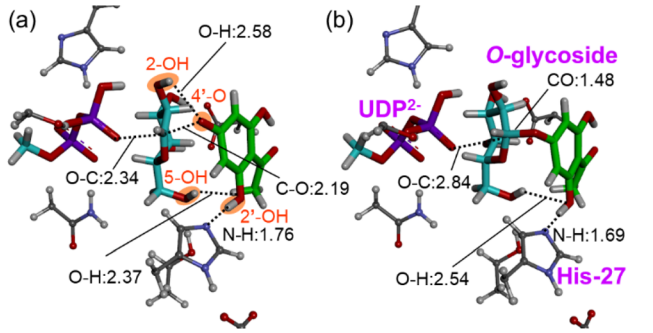



Fig. 8 Geometries around the reaction center in $TS_{1-6}$ (a) and **INT6** (b). Numbers in black and orange represent atomic distances in Å and position numbers shown in **INT0** in Fig. 3a, respectively. Atoms fixed during the geometry optimization were omitted. Their full views are also shown in Fig. S6.

aromatic character of the phloretin moiety. The instability of **INT6** could be due to the absence of proton transfer from the phloretin moiety to His-27, like **INT2** of the H27A mutant (see Fig. 6 and 8b). As mentioned above, both the 2′-*O*- and 4′-*O*-glycosylations had higher activation barriers than the *C*-glycosylation in Fig. 3, which was mainly due to the lack of H-bonds between **1a** and **2a**.

The mechanism of *O*-glycosylation obtained in this study, namely that protonation of the diphosphate leaving group promotes *O*-glycosylation, is similar to the previous studies on other *O*-glycosyltransferases.[27] Specifically, this proton has been reported to be supplied by a serine residue in LanGT2 (ref. 27*a*) and by a water molecule in RrUGT3.[27*b*] Our study of *C*-glycosylation reveals that the diphosphate is protonated *via* a relay from the aglycone substrate, which demonstrates that the protonation of the diphosphate moiety is a fundamentally important step in *C*-glycosylation, just as it is in *O*-glycosylation. Notably, whereas the aforementioned *O*-glycosyltransferases utilize an enzyme residue or a solvent molecule, the proton source for this *C*-glycosylation reaction is the substrate itself. This mechanistic divergence not only distinguishes *C*-glycosylation from *O*-glycosylation but also provides a basis for the enzyme's substrate selectivity.

Our proposed mechanism provides a plausible explanation for the experimentally observed selectivity toward certain substrates. For example, substrates containing a flopropione moiety, including phloretin, possess two potential reaction sites corresponding to those identified in Fig. 4 (*i.e.*, the 2′,3′,4′-positions and 6′,5′,4′-positions in **1a**). This structural feature enables initial *C*-glycosylation at the 3′-position, followed by a second *C*-glycosylation at the 5′-position, accounting for the formation of di-*C*-glycosides (see Fig. S7). Similarly, substrates containing a 2′,4′-dihydroxybenzene moiety offer only a single reactive site and therefore yield mono-*C*-glycosides, consistent with our mechanism. However, some substrates bearing the same reaction centers exhibit different selectivity.[14] This discrepancy is likely attributable to interactions between GgCGT and substrates far from the reaction center that are not captured in our current cluster model. To achieve a more comprehensive understanding of substrate scope and selectivity, calculations accounting for the overall GgCGT structure will be essential. Such studies are currently in progress and will be reported in due course.

## 4. Conclusions

We elucidated the reaction mechanism of GgCGT-catalyzed mono-*C*-glycosylation of phloretin with the coenzyme UDP-glucose. The most favorable reaction pathway starts with proton transfer from phloretin to the glucose moiety of UDP-glucose, followed by the $S_N2$ reaction, which involves the nucleophilic attack of phloretin on the glucose moiety and dissociation of UDP. This enzymatic reaction pathway afforded a non-aromatic intermediate, which was then promptly re-aromatized to produce a mono-*C*-glycoside, even in the absence of an enzymatic environment. The role of amino acids around the active site of GgCGT was also analyzed. The most


© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 12 August 2025. Downloaded on 1/21/2026 5:27:40 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

important residue, His-27, fixes the orientation of phloretin and stabilizes the TS of the rate-determining $S_N2$ reaction as well as the non-aromatic intermediate *via* proton abstraction from the phloretin moiety. Other residues mainly contributed to fixing the conformation of phloretin and UDP-glucose through an H-bond network. The role of His-27 in stabilizing the TS and the product of the $S_N2$ process was also confirmed by comparing the energy profiles of the same reaction catalyzed by GgCGT and its mutant, where His-27 was replaced by alanine. To understand the origin of the regioselectivity, *i.e.*, the selectivity for *C*- and *O*-glycosylation, the reaction pathway for *O*-glycosylation was examined. We revealed that proton transfer from phloretin to the glucose moiety also accelerated *O*-glycosylation, which differed from the previously proposed mechanism. The activation barrier for the *O*-glycosylation was 1.3 kcal $mol^{-1}$ higher than that for the *C*-glycosylation, mainly owing to the structural constraints by H-bond network among phloretin, glucose, and GgCGT. To the best of our knowledge, this is the first elucidation of a detailed mechanism for the enzymatic *C*-glycosylation. The finding that characteristic proton transfers are crucial has provided a new idea for enzyme design strategies.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

Comparison of the basis set, the dielectric constant (used for PCM), and the calculation model (cluster model versus full enzymatic system); full views of optimized geometries; details of EDA and IRC; summary of previously reported experimental results; Cartesian coordinates of all geometries. See DOI: **https://doi.org/10.1039/d5ra02643a**.

## Acknowledgements

This study was based on the results obtained from JSPS KAKENHI Grant No. JP20K05438, JP23H00288, JP24H01094, and JST Grant No. JPMJPF2221. We also acknowledge the computer resources provided by the Academic Center for Computing and Media Studies (ACCMS) at Kyoto University and the Research Center of Computer Science (RCCS) at the Institute for Molecular Science.

## References

1 (*a*) M. He, X. Zhou and X. Wang, *Signal Transduction Targeted Ther.*, 2024, **9**, 194; (*b*) G. W. Hart, *Curr. Opin. Cell Biol.*, 1992, **4**, 1017–1023; (*c*) S. Sirirungruang, C. R. Barnum, S. N. Tang and P. M. Shih, *Nat. Prod. Rep.*, 2023, **40**, 1170–1180; (*d*) C. J. Thibodeaux, C. E. Melançon, III and H. W. Liu, *Angew. Chem., Int. Ed.*, 2008, **47**, 9814–9859; (*e*) E. Kurze, M. Wüst, J. Liao, K. McGraphery, T. Hoffmann, C. Song and W. Schwab, *Nat. Prod. Rep.*, 2022, **39**, 389–409; (*f*) J. Ren, C. D. Barton and J. Zhan, *Biotechnol. Adv.*, 2023, **65**, 108146; (*g*) J. Yao, X. Xing, L. Yu, Y. Wang, X. Zhang and L. Zhang, *Ind. Crops Prod.*, 2022, **189**, 115784.

2 (*a*) Y. Q. Zhang, M. Zhang, Z. L. Wang, X. Qiao and M. Ye, *Biotechnol. Adv.*, 2022, **60**, 108030; (*b*) M. Pfeiffer and B. Nidetzky, *ACS Catal.*, 2023, **13**, 15910–15938; (*c*) E. Leclerc, X. Pannecoucke, M. Ethève-Quelquejeu and M. Sollogoub, *Chem. Soc. Rev.*, 2013, **42**, 4270–4283; (*d*) W. Zou, *Curr. Top. Med. Chem.*, 2005, **5**, 1363–1391; (*e*) P. Compain and O. R. Martin, *Bioorg. Med. Chem.*, 2001, **9**, 3077–3092.

3 Y. Sun, Z. Chen, J. Yang, I. Mutanda, S. Li, Q. Zhang, Y. Zhang, Y. Zhang and Y. Wang, *Commun. Biol.*, 2020, **3**, 110.

4 N. Putkaradze, D. Teze, F. Fredslund and D. H. Welner, *Nat. Prod. Rep.*, 2021, **38**, 432–443.

5 (*a*) G. Tegl and B. Nidetzky, *Biochem. Soc. Trans.*, 2020, **48**, 1583–1598; (*b*) T. Bililign, B. R. Griffith and J. S. Thorson, *Nat. Prod. Rep.*, 2005, **22**, 742–760; (*c*) A. Gutmann and B. Nidetzky, *Pure Appl. Chem.*, 2013, **85**, 1865–1877; (*d*) A. P. Rauter, R. G. Lopes and A. Martins, *Nat. Prod. Commun.*, 2007, **9**, 1175–1196; (*e*) M. A. Fischbach, H. Lin, D. R. Liu and C. T. Walsh, *Nat. Chem. Biol.*, 2006, **2**, 132–138; (*f*) M. A. Fischbach, H. Lin, D. R. Liu and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 571–576.

6 (*a*) K. Kitamura, Y. Ando, T. Matsumoto and K. Suzuki, *Chem. Rev.*, 2018, **118**, 1495–1598; (*b*) Y. Yang and B. Yu, *Chem. Rev.*, 2017, **117**, 12281–12356; (*c*) É. Bokor, S. Kun, D. Goyard, M. Tóth, J. P. Praly, S. Vidal and L. C. Somsák, *Chem. Rev.*, 2017, **117**, 1687–1764; (*d*) J. Stambaský, M. Hocek and P. Kocovský, *Chem. Rev.*, 2009, **109**, 6729–6764; (*e*) D. E. Kaelin Jr., O. D. Lopez and S. F. Martin, *J. Am. Chem. Soc.*, 2001, **123**, 6937–6938; (*f*) S. Hanessian and B. Lou, *Chem. Rev.*, 2000, **100**, 4443–4464.

7 (*a*) M. Li, Y. Zhou, Z. Wen, Q. Ni, Z. Zhou, Y. Liu, Q. Zhou, Z. Jia, B. Guo, Y. Ma, B. Chen, Z. M. Zhang and J. Wang, *Nat. Commun.*, 2024, **15**, 8893; (*b*) Y. Jiang, Y. Wei, Q. Zhou, G. Sun, X. Fu, N. Levin, Y. Zhang, W. Liu, N. Song, S. Mohammed, B. G. Davis and M. J. Koh, *Nature*, 2024, **631**, 320.

8 (*a*) X. Sheng and F. Himo, *Acc. Chem. Res.*, 2023, **56**, 938–947; (*b*) H. Y. Lin, X. Chen, J. Dong, J. F. Yang, H. Xiao, Y. Ye, L. H. Li, C. G. Zhan, W. C. Yang and G. F. Yang, *J. Am. Chem. Soc.*, 2021, **143**, 15674–15687; (*c*) K. Xie, X. Zhang, S. Sui, F. Ye and J. Dai, *Nat. Commun.*, 2020, **11**, 5162; (*d*) K. Steiner and H. Schwab, *Comput. Struct. Biotechnol. J.*, 2012, **2**, e201209010.

9 (*a*) D. Liang, J. Liu, H. Wu, B. Wang, H. Zhuabc and J. Qiao, *Chem. Soc. Rev.*, 2015, **44**, 8350–8374; (*b*) M. Krupička and I. Tvaroška, *J. Phys. Chem. B*, 2009, **113**, 11314–11319; (*c*) I. Tvaroška, S. Kozmon, M. Wimmerová and J. Koča, *J. Am. Chem. Soc.*, 2012, **134**, 15563–15571; (*d*) I. Tvaroška, S. Kozmon, M. Wimmerová and J. Koča, *Chem. - Eur. J.*, 2013, **19**, 8153–8162; (*e*) S. Kozmon and I. Tvaroška, *J. Am. Chem. Soc.*, 2006, **128**, 16921–16927.

10 C. Dürr, D. Hoffmeister, S. E. Wohlert, K. Ichinose, M. Weber, U. von Mulert, J. S. Thorson and A. Bechthold, *Angew. Chem., Int. Ed.*, 2004, **43**, 2962–2965.

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 12 August 2025. Downloaded on 1/21/2026 5:27:40 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

11 A. Gutmann and B. Nidetzky, *Pure Appl. Chem.*, 2013, **85**, 1865–1877.
12 Y. O. Bao, M. Zhang, X. Qiao and M. Ye, *Chem. Commun.*, 2022, **58**, 12337–12340.
13 (*a*) D. M. Liang, J. H. Liu, H. Wu, B. B. Wang, H. J. Zhu and J. J. Qiao, *Chem. Soc. Rev.*, 2015, **44**, 8350–8374; (*b*) T. Bililign, B. R. Griffith and J. S. Thorson, *Nat. Prod. Rep.*, 2005, **22**, 742–760; (*c*) J. Härle, S. Günther, B. Lauinger, M. Weber, B. Kammerer, D. L. Zechel, A. Luzhetskyy and A. Bechthold, *Chem. Biol.*, 2011, **18**, 520–530; (*d*) A. Gutmann and B. Nidetzky, *Angew. Chem., Int. Ed.*, 2012, **51**, 12879–12883; (*e*) L. Li, P. Wang and Y. Tang, *J. Antibiot.*, 2014, **67**, 65–70; (*f*) M. Liu, D. Wang, Y. Li, X. Li, G. Zong, S. Fei, X. Yang, J. Lin, X. Wang and Y. Shen, *Plant Cell*, 2020, **32**, 2917–2931.
14 M. Zhang, F. D. Li, K. Li, Z. L. Wang, Y. X. Wang, J. B. He, H. F. Su, Z. Y. Zhang, C. B. Chi, X. M. Shi, C. H. Yun, Z. Y. Zhang, Z. M. Liu, L. R. Zhang, D. H. Yang, M. Ma, X. Qiao and M. Ye, *J. Am. Chem. Soc.*, 2020, **142**, 3506–3512.
15 M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525–537.
16 X. Wang, *FEBS Lett.*, 2009, **583**, 3303–3309.
17 M. J. Abraham, D. van der Spoel, E. Lindahl and B. Hess, and the *GROMACS User Manual*, Version 5.1.5, GROMACS development team, 2017, **http://www.gromacs.org**.
18 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
19 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
20 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
21 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
22 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16*, Revision C.01, Gaussian, Inc., Wallingford CT, 2016.
23 T. Simonson and C. L. Brooks, III, *J. Am. Chem. Soc.*, 1996, **118**, 8452–8458.
24 K. Fukui, *Acc. Chem. Res.*, 1981, **14**, 363–368.
25 (*a*) S. Maeda, Y. Harabuchi, Y. Sumiya, M. Takagi, K. Suzuki, M. Hatanaka, Y. Osada, T. Taketsugu, K. Morokuma and K. Ohno, GRRM17, **http://iqce.jp/GRRM/index_e.shtml**; (*b*) S. Maeda, K. Ohno and K. Morokuma, *Phys. Chem. Chem. Phys.*, 2013, **15**, 3683–3701.
26 J. Liao, G. Sun, E. Kurze, W. Steinchen, T. D. Hoffmann, C. Song, Z. Zou, T. Hoffmann and W. G. Schwab, *Plant Commun.*, 2023, **4**, 100506.
27 (*a*) F. Mendoza and G. A. Jaña, *Org. Biomol. Chem.*, 2021, **19**, 5888; (*b*) M. Li, C. You, F. Guo, Q. Han, X. Xie, L. Ma, S. Li, L. Du, X. Sheng and H. Su, *Catal. Sci. Technol.*, 2024, **14**, 4882.


© 2025 The Author(s). Published by the Royal Society of Chemistry