



# Glycosylation Should Be Treated as Pharmacology Rather Than Decoration in Computational Models of Natural Product Exposure, Recognition, Transport, and Bioactivity

Patrick O'Connor<sup>1\*</sup>, Grace Murphy<sup>1</sup>, Niamh Walsh<sup>2</sup>

<sup>1</sup>Department of Glycosylation Pharmacology and Computational Modeling, School of Pharmacy, University College Cork, Cork, Ireland.

<sup>2</sup>Department of Recognition, Transport, and Bioactivity of Glycosylated Natural Products, Faculty of Pharmacy, University of Galway, Galway, Ireland.

## ABSTRACT

Natural-product glycosides are frequently represented computationally as scaffold variants whose sugar residues can be removed, simplified, or treated primarily as modifiers of bulk physicochemical properties. That simplification is sometimes reasonable, but it can also erase pharmacologically consequential information. This conceptual article examines glycosylation as a state variable that can alter molecular conformation, chemical stability, gastrointestinal transformation, epithelial transport, systemic chemical identity, cellular exposure, and target recognition. The central argument is not that glycosylation has a uniform directional effect, but that the sugar, linkage, substitution pattern, and transformation context may determine which molecular species should actually be modeled. A proposed glycosylation-aware pharmacology framework therefore treats an administered glycoside, its aglycone, conjugates, and microbially or enzymatically generated products as context-dependent members of an exposure–recognition hypothesis rather than interchangeable structures. The same logic extends to computation: topology-aware encodings, carbohydrate-sensitive force fields, explicit metabolite-state enumeration, and task-specific validation may be required when sugar chemistry governs the biological question. Conversely, aglycone-centered modeling remains defensible when presystemic conversion is dominant or the sugar does not materially alter exposure or recognition. The framework is intended to improve mechanistic discipline, not to establish a validated prediction rule. Its principal limitation is the heterogeneity of glycoside classes, transport systems, microbiomes, experimental models, and computational representations. Prospective matched glycoside–aglycone studies are required to determine when glycosylation materially changes prediction and lead interpretation.

**Key Words:** Natural products, Glycosides, Glycosylation, Pharmacology, Molecular modeling, Bioavailability

eIJPPR 2025; 15(5):111-119

**HOW TO CITE THIS ARTICLE:** O'Connor P, Murphy G, Walsh N. Glycosylation Should Be Treated as Pharmacology Rather Than Decoration in Computational Models of Natural Product Exposure, Recognition, Transport, and Bioactivity. Int J Pharm Phytopharmacol Res. 2025;15(5):111-9. <https://doi.org/10.51847/DcBljoO69G>

## INTRODUCTION

Glycosylation is often described as a structural modification, yet in natural-product pharmacology the attached sugar can participate in determining which chemical entity is presented to a biological system. Plant

glycosyltransferases generate extensive glycoside diversity, creating products that differ not only in sugar composition but also in linkage position and substitution pattern [1]. Large-scale analysis of natural-product databases likewise shows that glycosidic residues are structurally heterogeneous rather than a single

**Corresponding author:** Patrick O'Connor

**Address:** Department of Glycosylation Pharmacology and Computational Modeling, School of Pharmacy, University College Cork, Cork, Ireland.

**E-mail:** [patrick.oconnor@ucc.ie](mailto:patrick.oconnor@ucc.ie)

**Received:** 04 June 2025; **Revised:** 15 August 2025; **Accepted:** 16 August 2025

This is an **open access** journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.



interchangeable appendage [2]. This diversity matters because the nominal compound entered into a computational workflow may not correspond to the molecular species that dominates absorption, metabolism, circulation, cellular entry, or target recognition.

Analytical discovery itself illustrates the problem. Tailored integration of biosynthetic information with mass-spectrometric metabolomics can reveal glycosylated natural products that would be difficult to interpret if sugar-bearing forms were treated merely as redundant versions of their aglycones [3]. Chemical-state identification is therefore upstream of pharmacological modeling: an incorrect assumption about whether the relevant entity is an intact glycoside, a hydrolyzed aglycone, or a downstream metabolite can distort every subsequent inference.

The position advanced here is deliberately conditional. Glycosylation does not impose one universal change in permeability, stability, potency, or exposure. In some compounds, the sugar may be pharmacologically dominant; in others, it may mainly regulate storage, solubility, processing, or delivery before being removed. The appropriate question is therefore not whether glycosides are generally “better” or “worse” drug-like molecules. It is which molecular state carries the pharmacologically relevant information for the specific biological process being modeled. Treating glycosylation as pharmacology means preserving that question throughout chemical representation, exposure modeling, mechanistic interpretation, and validation rather than deciding in advance that the aglycone is the only meaningful structure.

#### *Why computational pipelines often treat glycosides as minor variants*

Many molecular pipelines are optimized around compact drug-like structures, standardized graphs, canonicalized

chemical states, and descriptors that compress chemistry into representations suitable for similarity search, property prediction, docking, or machine learning. Within such workflows, a sugar can appear to be an inconvenient peripheral group: large, stereochemically dense, conformationally flexible, and frequently associated with high polarity. Simplification is consequently understandable, but its pharmacological validity must be demonstrated rather than assumed.

Conformational analysis of flavonoid glycosides shows that sugar-linked chemistry can alter accessible molecular arrangements, meaning that an aglycone pose or conformational ensemble is not necessarily transferable unchanged to its glycoside [4]. At the representation level, GlyLES was developed because glycan connectivity, stereochemistry, and linkage information require explicit parsing before they can be translated reliably into machine-readable structures such as SMILES [5]. Graph-based learning provides a complementary demonstration: glycan topology contains structured information that can be learned for predictive tasks when represented explicitly rather than flattened into coarse composition alone [6].

The conceptual error, therefore, is not simplification itself. It is the untested collapse of chemically distinct states that may support different exposure or recognition hypotheses. An aglycone-only representation can remain defensible when it corresponds to the process actually being modeled—for example, when dominant conversion makes the aglycone the relevant biological species. Where that condition is unknown, however, glycoside removal is not neutral preprocessing; it is a pharmacological assumption. The constructs generated by this distinction are consolidated in **Table 1**, which separates the chemical information being simplified from the assumption introduced when that information is removed.

**Table 1.** Chemical-state constructs that determine whether glycoside simplification is pharmacologically defensible

| Construct                      | Working definition                                                             | Computational consequence                                                                         | Evidence status and boundary                                                                                                                         |
|--------------------------------|--------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Glycosylation state</b>     | Identity of the intact sugar-bearing molecular form under consideration        | Preserve the observed parent structure before deciding whether an aglycone surrogate is justified | Natural-product glycosides show substantial structural heterogeneity [2]. Structural occurrence alone does not establish pharmacological importance. |
| <b>Linkage topology</b>        | Position, stereochemistry, and connectivity joining sugar and scaffold         | Encode linkage explicitly when transformation, geometry, or molecular identity depends on it      | Dedicated glycan parsing preserves connectivity and stereochemical information [5].                                                                  |
| <b>Molecular presentation</b>  | Accessible three-dimensional arrangements produced by scaffold–sugar chemistry | Avoid transferring an aglycone conformational ensemble automatically to the glycoside             | Glycoside chemistry can change modeled conformational space [4]. Computational differences do not themselves prove biological consequences.          |
| <b>Representation collapse</b> | Replacement of the glycoside by its aglycone                                   | Treat removal as a task-specific modeling assumption                                              | Graph learning demonstrates that glycan topology contains usable predictive information                                                              |

or by a coarse sugar  
descriptor

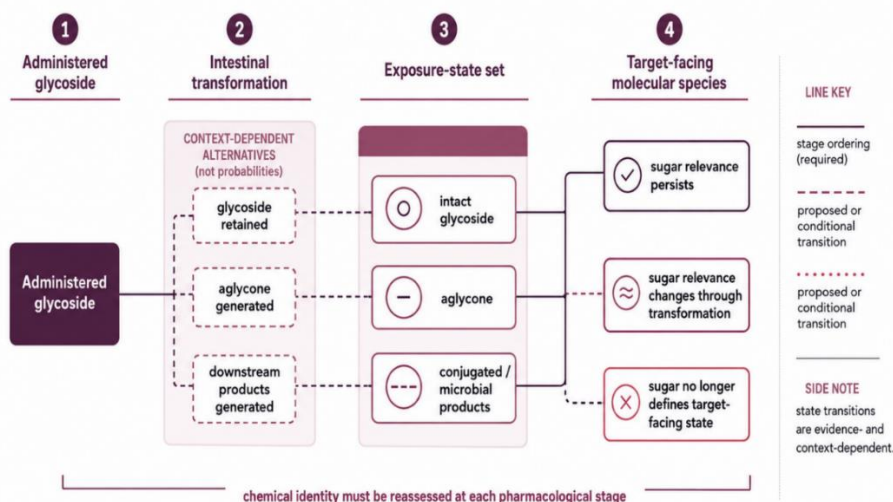
rather than routine  
normalization

[6]. This does not validate glycoside-aware  
natural-product lead ranking.

The same argument is better understood relationally than as a list of descriptors. **Figure 1** therefore represents glycosylation state as the point at which observed chemical

identity, computational representation, and the biological hypothesis being modeled must remain aligned.

#### Conditional persistence of glycosylation across transformation, exposure, and target recognition



**Figure 1.** From glycosylated chemical identity to a defensible computational pharmacology hypothesis

#### *Glycosylation changes solubility and molecular presentation*

The most familiar intuition about glycosylation is physicochemical: adding a sugar can change polarity, hydration, and apparent solubility. For pharmacological modeling, however, bulk polarity is only one component of molecular presentation. Sugar attachment also changes hydrogen-bonding capacity, steric shape, intramolecular interactions, conformational preferences, and susceptibility to environmental conditions. These effects need not move in the same direction, and they should not be reduced to a universal formulation that glycosylation simply makes a molecule “more soluble but less permeable.”

A useful experimental illustration comes from fidaxomicin, a complex glycosylated natural product. Formal single-atom editing of this scaffold altered acid stability while retaining antibacterial activity in the studied analogue context [7]. This finding neither shows that glycosylation universally improves stability nor isolates every contribution made by a sugar residue. It demonstrates a more consequential principle for computation: closely related structures within a glycosylated architecture can separate chemical stability from functional activity, so structural similarity alone does not ensure an equivalent developability profile.

Conformation adds another level. Glycoside-specific molecular modeling indicates that sugar-linked chemistry can shift accessible conformational states [4]. A workflow that estimates physicochemical properties, docking poses, membrane behavior, or intramolecular hydrogen bonding from a single aglycone-like geometry may therefore discard a property-relevant ensemble before prediction begins. The appropriate strategy is not to assign a predetermined favorable or unfavorable effect to glycosylation, but to identify which aspects of molecular presentation are sugar-sensitive for the compound and endpoint under study. Descriptor selection, conformational sampling, and experimental corroboration can then be matched to that uncertainty. Directional claims about solubility, stability, or presentation should remain compound-specific unless measured directly.

#### *Membrane permeation and transporter recognition*

Glycosides are frequently expected to permeate biological membranes poorly because sugar attachment often increases molecular size and polarity. That expectation can function as a prior, but it is not a pharmacological rule. The anthocyanin absorption literature shows substantial structural and mechanistic heterogeneity, including evidence that intact glycosylated species can participate in intestinal uptake [8]. Direct mass-spectrometric visualization in rats has further demonstrated intestinal absorption-associated signals for defined acylated

anthocyanins, showing that intact or still-glycosylated molecular states cannot always be excluded before exposure is investigated [9].

Transporter biology further complicates passive-permeability reasoning. Evidence on dietary flavonoids indicates that epithelial transport proteins can contribute to transepithelial handling alongside passive diffusion, local metabolism, and efflux [10]. Consequently, a prediction of low passive permeability does not by itself establish low cellular or systemic exposure. A polar glycoside may interact with transport machinery, may be transformed before membrane passage, or may generate products with substantially different permeability characteristics.

Computational pharmacology should therefore represent these possibilities as competing exposure hypotheses rather than compress them into one scalar concept of “drug-likeness.” Passive-permeability prediction describes one route. Transporter recognition addresses another. Presystemic transformation changes the identity of the species to which either prediction applies. Confounding these levels can make a numerically plausible prediction mechanistically misleading. Sugar presence should thus prompt questions about transport mechanism and persistence of chemical state rather than an automatic permeability penalty or advantage. Transporter-specific interpretations must remain bounded by the compound classes, biological systems, and experimental conditions in which those mechanisms were observed; neither intact glycoside uptake nor transporter involvement should be presumed for natural-product glycosides as a class.

#### *Intestinal hydrolysis and microbial deglycosylation*

The pharmacological identity of an orally administered glycoside may change before systemic exposure. This is not captured adequately by a generic “hydrolysis” flag because cleavage depends on linkage chemistry, enzyme availability, and microbial context. C-glycosides illustrate the problem particularly clearly. Specialized microbial enzymes capable of cleaving carbon–carbon glycosidic bonds have been identified and structurally characterized, demonstrating that C-glycosides need not follow the transformation logic used for conventional O-glycosides [11]. Enzymatic  $\beta$ -elimination provides another route for deglycosylation in defined O- and C-glycoside systems [12], while glycoside 3-oxidases reveal oxidative chemistry associated with C-glycoside metabolism [13]. These mechanisms support a computational distinction between “sugar present in the administered structure” and “sugar retained at the biological site being modeled.” C-glycoside cleavage can involve specialized enzymatic chemistry and therefore cannot be inferred from generic O-glycoside hydrolysis rules [11–13]. Microbial fermentation adds further uncertainty: in vitro human fecal systems

demonstrate that flavonoids can be transformed within a changing microbial environment [14]. Such findings establish metabolic possibilities, not universal human conversion rates. A glycosylation-aware model should therefore treat linkage class, enzyme context, and microbiome-dependent transformation as conditional state transitions, with uncertainty retained when the relevant pathway has not been measured.

#### *Systemic exposure to glycosides, aglycones, and downstream metabolites*

Once transformation is admitted, “exposure to the compound” becomes chemically ambiguous. Human quercetin evidence is instructive because bioavailability varies with the administered form and formulation context [15]. The relevant systemic species may also include phase-II conjugates whose concentrations and properties differ from those of the nominal aglycone; critical evaluation of human data warns against attributing biological activity automatically to the parent structure used in many in vitro experiments [16].

The computational consequence is an exposure-state set rather than a single canonical molecule. For a given natural product, that set may contain intact glycoside, aglycone, host conjugates, and microbial products, but membership should be evidence-driven rather than exhaustively invented. Each candidate state can then be linked to the process for which it is plausible: intestinal transport, circulating exposure, tissue entry, or target engagement. This proposed representation does not imply that every state is abundant or active. Its purpose is to prevent a model from assigning a prediction to one structure while the experimental or physiological observation belongs predominantly to another. Exposure modeling should therefore preserve chemical provenance and explicitly state when the pharmacologically relevant species remains unresolved.

#### *Direct target recognition by glycosylated natural products*

Glycosylation can matter after exposure as well as before it. Structural work on cardiotonic steroids bound to Na<sup>+</sup>,K<sup>+</sup>-ATPase provides a direct example in which the architecture of a glycosylated natural product participates in a defined ligand–target recognition system [17]. This evidence is valuable precisely because it is narrow: it establishes that sugar-containing architecture can contribute to molecular recognition, not that sugar residues universally improve affinity or that every glycoside requires intact sugar for activity.

For computational pharmacology, target recognition should therefore be tested as a separate question from bioavailability. An aglycone may be the dominant circulating species yet bind differently from the

administered glycoside; alternatively, an intact glycoside may reach a target for which its sugar contributes steric or interaction features. Docking or free-energy calculations that remove the sugar before establishing which species reaches the target risk answering the wrong molecular

question. Conversely, retaining a sugar without credible exposure is equally problematic.

**Table 2** distinguishes these levels so that “glycosylation matters” is not used as a single undifferentiated mechanistic claim.

**Table 2.** Distinguishing pharmacological roles of the sugar across exposure and target-recognition hypotheses

| Role of sugar                     | Question addressed                                                                   | Evidence-compatible interpretation                                                                                        | What must not be inferred                                                   |
|-----------------------------------|--------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>Transformation determinant</b> | Is the glycosidic linkage retained, cleaved, or chemically modified before exposure? | Specialized C-glycoside cleavage makes linkage context mechanistically relevant [11].                                     | Cleavage occurs universally or at equivalent rates across hosts.            |
| <b>Exposure determinant</b>       | Which chemical species reaches systemic circulation?                                 | Human quercetin evidence supports form-dependent bioavailability and downstream conjugation [15].                         | The administered glycoside is necessarily the dominant circulating species. |
| <b>Recognition determinant</b>    | Does intact glycosylated architecture participate in target binding?                 | Defined cardiotonic-steroid structures show that a glycosylated ligand can function as an intact recognition entity [17]. | Sugar-mediated recognition generalizes across glycoside classes.            |
| <b>Proposed integrative role</b>  | Which state should a computational model represent for a specific endpoint?          | Match molecular representation to the evidenced transformation, exposure, and recognition state.                          | One representation is sufficient for every pharmacological endpoint.        |

#### *Glycosylation and cellular uptake*

Cellular uptake sits between systemic exposure and intracellular pharmacology, but it should not be collapsed into passive membrane permeability. Evidence for anthocyanin absorption and flavonoid transport indicates that intact glycoside handling can involve multiple processes, including transporter participation, local metabolism, and movement of chemically transformed species [9]. Transport-protein literature further supports a context-dependent role for epithelial uptake and efflux systems [10].

A useful modeling discipline is therefore to separate four questions: whether the intact glycoside encounters the cell, whether it can cross the membrane passively, whether transport machinery recognizes it, and whether extracellular or intracellular conversion changes the entering or retained species. These questions can yield different answers for the same parent natural product. A low passive-permeability estimate should not be reinterpreted as proof of absent cellular exposure, while transporter plausibility should not be treated as evidence of uptake without a compatible biological system. The proposed framework treats cellular uptake as a state-transition problem constrained by measured transport, localization, and metabolism whenever available. Where such data are absent, predictions should retain uncertainty over the identity of the species that actually enters the cell.

#### *Why aglycone-only modeling can produce the wrong mechanistic story*

Aglycone-only modeling can fail before pharmacology is even considered. In mass-spectrometric identification,

computational modification of an aglycone spectral library can improve matching of glycosides, demonstrating that parent–sugar relationships carry analytically useful information [18]. The pharmacological analogy is conceptual rather than evidentiary: when glycoside and aglycone are collapsed prematurely, the model may preserve the scaffold while losing the state information needed to interpret a measurement.

Three distinct mismatches can result. A representation mismatch occurs when the modeled structure is not the experimentally observed chemical species. An exposure mismatch occurs when predictions are assigned to the administered molecule although metabolism produces other dominant species. A recognition mismatch occurs when binding is calculated for an aglycone despite evidence or plausible uncertainty that an intact glycoside reaches the target. These categories are proposed here as diagnostic distinctions, not validated error classes.

Aglycone-only modeling is therefore not intrinsically wrong. It is wrong when the aglycone assumption precedes, rather than follows, evidence about transformation and exposure. In contexts with rapid, dominant deglycosylation and negligible intact exposure, the aglycone may be exactly the appropriate computational entity. The methodological requirement is to justify the collapse and specify the endpoint to which it applies.

#### *A glycosylation-aware pharmacology framework*

The proposed framework organizes glycosylation-aware modeling around chemical state rather than around an a priori preference for either glycoside or aglycone. Its first element is identity: define the observed glycoside, sugar

composition, linkage, attachment site, and relevant stereochemistry. Its second is transformation: enumerate only experimentally plausible transitions. High-throughput profiling of plant family-1 glycosyltransferases demonstrates extensive enzyme–substrate specificity, reinforcing that glycosylation outcomes depend on biochemical context [19]. Rational tuning of glycosyltransferase chemoselectivity further shows that linkage formation is an experimentally manipulable molecular variable rather than a generic decoration [20].

The third element is biological processing. Engineering a yeast chassis by removing endogenous glycosidases can change production of glycosylated drug-like natural products, illustrating that cellular enzyme background influences which glycosylation states persist [21]. This production system should not be treated as a model of human metabolism; it is evidence for the broader principle that biological context determines chemical-state persistence. Chemical and biocatalytic glycosylation strategies likewise involve different constraints, underscoring that “a glycoside” is not one mechanistically uniform construction [22].

The final elements are exposure and recognition. For each endpoint, the model should identify which chemical state is supported, which alternatives remain plausible, and where uncertainty enters. The framework therefore does not prescribe that every glycoside be docked, simulated, and pharmacokinetically modeled in all possible forms. It proposes a state-selection discipline: retain complexity only where evidence indicates that sugar chemistry can alter the biological question, and simplify only when the resulting assumption is explicit and testable.

A practical consequence is that models should expose the state-selection decision rather than bury it in preprocessing. If an intact glycoside is retained for docking but an aglycone is used for permeability prediction, the workflow should state why those endpoints justify different molecular entities. If several states remain

plausible, predictions should be reported as conditional on state rather than averaged into a chemically ambiguous score. This makes disagreement between computational modules diagnostically useful: it can reveal where transformation, exposure, or recognition evidence is missing.

#### *Representing glycosides in molecular prediction systems*

Carbohydrates impose computational challenges that are not captured fully by adding a few conventional descriptors. Computational glycoscience emphasizes their conformational diversity, stereochemical density, solvent interactions, and multiscale behavior [23]. Direct comparison of carbohydrate force fields in protein–carbohydrate complexes shows that simulation results can depend materially on the chosen parameterization [24]. Coarse-grained approaches similarly require carbohydrate-specific models; the Martini 3 carbohydrate force field was developed to reproduce selected structural and interaction behavior at its intended resolution [25]. Carbohydrate simulation is therefore representation- and force-field-sensitive, requiring explicit parameterization and validation appropriate to resolution and task [23–25].

Representation also matters beyond molecular dynamics. DeepGlycanSite shows that specialized learning architectures can predict carbohydrate-binding sites from protein structural information [26], while deep learning applied to tandem mass spectrometry can recover glycan structural information from large spectral datasets [27]. Neither study validates natural-product glycoside bioactivity prediction, but together they demonstrate that sugar-specific information can be encoded and learned rather than discarded. Recent systematic parametrization of disaccharides for Martini 3 further illustrates the importance of residue- and linkage-aware validation [28]. The manuscript-derived computational dimensions are summarized in **Table 3**. They are proposed as minimum questions for model design, not as a validated checklist.

**Table 3.** Proposed dimensions for glycoside-aware molecular representation and model validation

| Dimension                            | Modeling question                                                               | Evidence basis                                                                                            | Proposed practice                                                           |
|--------------------------------------|---------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>Chemical identity</b>             | Which glycoside, linkage, stereochemistry, and attachment state is represented? | Glycan parsing requires explicit structural semantics [5].                                                | Preserve the experimentally relevant chemical state before normalization.   |
| <b>Conformational representation</b> | Does sugar attachment alter the accessible molecular ensemble?                  | Glycoside modeling supports conformational sensitivity [4].                                               | Sample sugar-bearing conformations when the endpoint is geometry-sensitive. |
| <b>Interaction model</b>             | Are carbohydrate interactions parameterized for the intended resolution?        | Force-field benchmarks and carbohydrate-specific models show parameter dependence [24, 25].               | Report force field and test sensitivity for critical interactions.          |
| <b>Learned representation</b>        | Can topology or carbohydrate-recognition information be retained?               | Glycan graph learning and binding-site prediction demonstrate learnable sugar-specific structure [6, 26]. | Avoid stripping topology when it is plausibly predictive.                   |

|                            |                                                                               |                                                                                                 |                                                                     |
|----------------------------|-------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|
| <b>Analytical linkage</b>  | Can modeled identity be reconciled with observed glycoside evidence?          | Glycan MS learning and glycoside spectral matching retain sugar-related information [18, 27].   | Link predicted states to experimentally distinguishable species.    |
| <b>Validation boundary</b> | Does performance persist across residue, linkage, and chemical-state changes? | Systematic disaccharide parametrization illustrates residue- and linkage-level validation [28]. | Test matched glycoside–aglycone or linkage variants where feasible. |

#### *Contexts in which the sugar is pharmacologically dominant*

“Pharmacologically dominant” should not mean that a sugar alone causes activity. It describes contexts in which changing or removing the sugar changes the identity of the hypothesis being tested. Dominance is most plausible when sugar chemistry controls a transformation boundary, a transport interaction, a stability-relevant molecular presentation, or a direct recognition feature. Fidaxomicin illustrates sensitivity of a glycosylated scaffold to stability-related structural editing [7]; specialized C-glycoside metabolism demonstrates linkage-dependent transformation [11]; and cardiotonic-steroid structures establish a case in which intact glycosylated architecture participates directly in target recognition [17].

The converse is equally important. If a glycoside is rapidly and reproducibly converted to an aglycone before the modeled endpoint, if intact exposure is negligible, and if downstream activity is attributable to the converted species, then the sugar may be pharmacologically important as an exposure determinant without needing to remain in the target-binding model. Dominance is therefore endpoint-specific. The same sugar can be decisive for intestinal processing, irrelevant to systemic target recognition, and still important to formulation or stability. This prevents the framework from replacing an aglycone bias with an equally unjustified glycoside bias.

#### *Boundary conditions and experimental priorities*

The framework is bounded first by chemical heterogeneity. Intact absorption and transporter involvement are compound- and context-specific, so sugar presence alone cannot determine permeability [8–10]. Linkage chemistry likewise prevents one deglycosylation rule from being transferred across O- and C-glycosides. Second, computational conclusions depend on representation quality: force-field performance in protein–carbohydrate systems can vary with parameterization [24]. Third, systemic observations may belong to conjugated or transformed species rather than the nominal parent compound [16].

Prospective validation should therefore compare matched molecular states rather than merely compare algorithms. Useful designs include paired glycoside–aglycone measurements, linkage variants where chemically feasible, transporter perturbation, gastrointestinal or microbial

transformation assays, and analytical discrimination of parent and downstream products. For computational models, ablation can test whether retaining sugar topology improves or worsens the endpoint under a fixed validation design. Structural or target-engagement experiments are needed where the claim concerns recognition rather than exposure.

Validation should also preserve endpoint separation. A permeability experiment cannot by itself validate a target-binding representation, and a docking result cannot establish which molecular species reaches the target. Concordance becomes informative only when chemical identity is tracked across the relevant sequence of measurements. Where matched experiments are impractical, the model should identify the unsupported transition explicitly rather than converting it into an assumed mechanistic link. This preserves falsifiability and prevents uncertainty in one pharmacological layer from being hidden by confidence in another.

No universal threshold should decide when glycosylation “matters.” The proposed criterion is evidential: a sugar deserves explicit pharmacological representation when plausible changes in its identity or persistence could alter the species, process, or endpoint to which the prediction is attributed.

## CONCLUSION

Glycosylation should be treated neither as an automatic advantage nor as dispensable decoration. Its pharmacological importance arises from the possibility that sugar composition, linkage, persistence, and transformation change which molecular species is exposed to a biological system and which structure participates in recognition. The central contribution of this article is therefore a proposed state-based framework: identify the glycosylated chemical entity, determine plausible transformations, distinguish exposure from recognition, select the molecular representation appropriate to the endpoint, and preserve uncertainty where the relevant state is unresolved.

This framework does not establish that glycoside-aware models will outperform aglycone-centered models, nor does it prescribe universal retention of every sugar. In some systems, the aglycone remains the correct pharmacological entity. The practical requirement is

narrower and more defensible: simplification should follow biological justification rather than precede it. Matched experimental and computational studies are needed to determine when sugar-aware representation changes prediction, mechanism, or lead interpretation. Until that evidence exists, the glycan should remain part of the drug hypothesis whenever removing it would silently change the pharmacological question.

**Acknowledgments:** None

**Conflict of interest:** None

**Financial support:** None

**Ethics statement:** None

## REFERENCES

- [1] Sirirungruang S, Barnum CR, Tang SN, Shih PM. Plant glycosyltransferases for expanding bioactive glycoside diversity. *Nat Prod Rep.* 2023;40(7):1170-80. doi:10.1039/D2NP00077F
- [2] Schaub J, Zielesny A, Steinbeck C, Sorokina M. Description and analysis of glycosidic residues in the largest open natural products database. *Biomolecules.* 2021;11(4):486. doi:10.3390/biom11040486
- [3] Chen D, Song Z, Han J, Liu J, Liu H, Dai J. Targeted discovery of glycosylated natural products by tailoring enzyme-guided genome mining and MS-based metabolome analysis. *J Am Chem Soc.* 2024;146(14):9614-22. doi:10.1021/jacs.3c12895
- [4] Alkhodier RA, Mishra SK, Doerksen RJ, Colby DA. Comparison of conformational analyses of naturally occurring flavonoid-O-glycosides with unnatural flavonoid-CF<sub>2</sub>-glycosides using molecular modeling. *J Chem Inf Model.* 2023;63(2):375-86. doi:10.1021/acs.jcim.2c01147
- [5] Joeres R, Bojar D, Kalinina OV. GlyLES: Grammar-based parsing of glycans from IUPAC-condensed to SMILES. *J Cheminform.* 2023;15(1):37. doi:10.1186/s13321-023-00704-0
- [6] Burkholz R, Quackenbush J, Bojar D. Using graph convolutional neural networks to learn a representation for glycans. *Cell Rep.* 2021;35(11):109251. doi:10.1016/j.celrep.2021.109251
- [7] Ferrara I, Chesnokov GA, Dittmann S, Blacque O, Sievers S, Gademann K. Formal single atom editing of the glycosylated natural product fidaxomicin improves acid stability and retains antibiotic activity. *JACS Au.* 2024;4(6):2267-80. doi:10.1021/jacsau.4c00206
- [8] Hahm TH, Tanaka M, Matsui T. Current knowledge on intestinal absorption of anthocyanins. *J Agric Food Chem.* 2022;70(8):2501-9. doi:10.1021/acs.jafc.1c08207
- [9] Hahm TH, Tanaka M, Nguyen HN, Tsutsumi A, Aizawa K, Matsui T. Matrix-assisted laser desorption/ionization mass spectrometry-guided visualization analysis of intestinal absorption of acylated anthocyanins in Sprague-Dawley rats. *Food Chem.* 2021;334:127586. doi:10.1016/j.foodchem.2020.127586
- [10] Li S, Liu Y. Intestinal absorption mechanism and nutritional synergy promotion strategy of dietary flavonoids: Transintestinal epithelial pathway mediated by intestinal transport proteins. *Crit Rev Food Sci Nutr.* 2025;65(22):4368-81. doi:10.1080/10408398.2024.2387320
- [11] Mori T, Kumano T, He H, Watanabe S, Senda M, Moriya T, et al. C-Glycoside metabolism in the gut and in nature: Identification, characterization, structural analyses and distribution of C-C bond-cleaving enzymes. *Nat Commun.* 2021;12(1):6294. doi:10.1038/s41467-021-26585-1
- [12] Bitter J, Pfeiffer M, Borg AJE, Kuhlmann K, Pavkov-Keller T, Sánchez-Murcia PA, et al. Enzymatic  $\beta$ -elimination in natural product O- and C-glycoside deglycosylation. *Nat Commun.* 2023;14(1):7123. doi:10.1038/s41467-023-42750-0
- [13] Taborda A, Frazão T, Rodrigues MV, Fernández-Luengo X, Sancho F, Lucas MF, et al. Mechanistic insights into glycoside 3-oxidases involved in C-glycoside metabolism in soil microorganisms. *Nat Commun.* 2023;14(1):7289. doi:10.1038/s41467-023-42000-3
- [14] Pan L, Ye H, Pi X, Liu W, Wang Z, Zhang Y, et al. Effects of several flavonoids on human gut microbiota and its metabolism by in vitro simulated fermentation. *Front Microbiol.* 2023;14:1092729. doi:10.3389/fmicb.2023.1092729
- [15] Liu L, Barber E, Kellow NJ, Williamson G. Improving quercetin bioavailability: A systematic review and meta-analysis of human intervention studies. *Food Chem.* 2025;477:143630. doi:10.1016/j.foodchem.2025.143630
- [16] Williamson G, Clifford MN. A critical examination of human data for the biological activity of quercetin and its phase-2 conjugates. *Crit Rev Food Sci Nutr.* 2025;65(9):1669-705. doi:10.1080/10408398.2023.2299329
- [17] Kanai R, Cornelius F, Ogawa H, Motoyama K, Vilsen B, Toyoshima C. Binding of cardiotonic steroids to Na<sup>+</sup>,K<sup>+</sup>-ATPase in the E2P state. *Proc Natl Acad Sci U S A.* 2021;118(1). doi:10.1073/pnas.2020438118

- [18] Rodríguez EP, Li Y, Vaniya A, Shih PM, Fiehn O. Alternative identification of glycosides using MS/MS matching with an in silico-modified aglycone mass spectra library. *Anal Chem.* 2023;95(28):10618-24. doi:10.1021/acs.analchem.3c00957
- [19] Sirirungruang S, Blay V, Rodriguez EP, Scott YF, Vuu KM, Barnum CR, et al. A substrate-multiplexed platform for profiling enzymatic potential of plant family 1 glycosyltransferases. *Nat Commun.* 2025;16(1):6366. doi:10.1038/s41467-025-61530-6
- [20] Li M, Zhou Y, Wen Z, Ni Q, Zhou Z, Liu Y, et al. An efficient C-glycoside production platform enabled by rationally tuning the chemoselectivity of glycosyltransferases. *Nat Commun.* 2024;15(1):8893. doi:10.1038/s41467-024-53209-1
- [21] Huang Y, Zhong W, Varga KE, Benkő Z, Pócsi I, Yang C, et al. Promoting the glycosylation of drug-like natural products in a *Saccharomyces cerevisiae* chassis by deletion of endogenous glycosidases. *Bioresour Technol.* 2025;422:132258. doi:10.1016/j.biortech.2025.132258
- [22] Mejia-Otalvaro F, Lax BM, Kirtel O, Welner DH. Sustainable natural product glycosylation: A critical evaluation of biocatalytic and chemical approaches. *ChemSusChem.* 2025;18(20). doi:10.1002/cssc.202501094
- [23] Perez S, Makshakova O. Multifaceted computational modeling in glycoscience. *Chem Rev.* 2022;122(20):15914-70. doi:10.1021/acs.chemrev.2c00060
- [24] Płazińska A, Płaziński W. Comparison of carbohydrate force fields in molecular dynamics simulations of protein-carbohydrate complexes. *J Chem Theory Comput.* 2021;17(4):2575-85. doi:10.1021/acs.jctc.1c00071
- [25] Grünwald F, Punt MH, Jefferys EE, Vainikka PA, König M, Virtanen V, et al. Martini 3 coarse-grained force field for carbohydrates. *J Chem Theory Comput.* 2022;18(12):7555-69. doi:10.1021/acs.jctc.2c00757
- [26] He X, Zhao L, Tian Y, Li R, Chu Q, Gu Z, et al. Highly accurate carbohydrate-binding site prediction with DeepGlycanSite. *Nat Commun.* 2024;15(1):5163. doi:10.1038/s41467-024-49516-2
- [27] Urban J, Jin C, Thomsson KA, Karlsson NG, Ives CM, Fadda E, et al. Predicting glycan structure from tandem mass spectrometry via deep learning. *Nat Methods.* 2024;21(7):1206-15. doi:10.1038/s41592-024-02314-6
- [28] Brandner AF, Smith IPS, Marrink SJ, Souza PCT, Khalid S. Systematic approach to parametrization of disaccharides for the Martini 3 coarse-grained force field. *J Chem Inf Model.* 2025;65(3):1537-48. doi:10.1021/acs.jcim.4c01874