INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES
[ISSN: 0975-4725; CODEN(USA): IJPS00]
Journal Homepage: https://www.ijpsjournal.com

OPEN ACCESS

Research Article

A Focused Computational Study of *Clitoria ternatea* Constituents Against GAL3ST1 for a Substrate-Reduction Strategy in MLD.

R. Elakkiya*, R. Praveenkumar, R. Riyaskhan, I. Vinnarivu

Department of Pharmaceutical Chemistry,Sri Rangapoopathi college of Pharmacy,Alampoondi,villupuram-604151.

ARTICLE INFO

Published: 26 Nov 2026

Keywords:
cerebroside sulfotransferase, GAL3ST1, sulfatide, blind molecular docking, CB-Dock2, quercetin, kaempferol, isorhamnetin, ADMET, ProTox-3.0, SwissTargetPrediction, flavonol, in silico drug discovery, metachromatic leukodystrophy, substrate-reduction therapy, Clitoria ternatea, butterfly pea flower

DOI:
10.5281/zenodo.24548653

ABSTRACT

Background: Cerebroside sulfotransferase (GAL3ST1), also known as galactosylceramide sulfotransferase or CST, is a 423-residue Golgi-resident type II transmembrane enzyme that catalyses the terminal, sulfate-transfer step in sulfatide biosynthesis. Elevated GAL3ST1 expression and activity have been linked to renal cell carcinoma, ovarian carcinoma, and altered insulin sensitivity, yet the protein has no experimentally resolved crystal structure and no formally annotated ligand-binding domain, which has limited structure-based drug discovery against it. Objective: To evaluate whether three structurally related dietary flavonols — quercetin, kaempferol, and isorhamnetin, present together in Clitoria ternatea (butterfly pea flower) — can occupy a druggable cavity on a computational model of human GAL3ST1, and to characterise their predicted pharmacokinetic, toxicological, and off-target profiles. A secondary, hypothesis-generating aim was to explore, purely at the computational level, whether this low-cost, dietarily available flavonol source could offer a rational, substrate-reduction-based starting point for eventual research relevant to metachromatic leukodystrophy (MLD) — a disease whose currently available disease-modifying therapies are among the costliest in medicine — without at any point claiming that GAL3ST1 inhibition, blood-brain-barrier penetration, or therapeutic benefit in MLD has been established. Methods: Because no validated active-site annotation exists for GAL3ST1, an unbiased, cavity-detection-guided blind docking strategy was adopted using CB-Dock2, which combines curvature-based cavity detection (CurPocket) with AutoDock Vina scoring. Five candidate cavities were identified on the modelled receptor and each flavonol was docked into all five..

***Corresponding Author:** R.ELAKKIYA
Address: Department of Pharmaceutical Chemistry,Sri Rangapoopathi college of Pharmacy,Alampoondi,villupuram-604151.
Email ✉: elakkiyaravi93@gmail.com

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Drug-likeness and pharmacokinetic behaviour were assessed with SwissADME, acute and organ-level toxicity with Pro-Tox-3.0, and putative off-target interactions with SwissTargetPrediction

INTRODUCTION

1.1 Glycosphingolipids and Sulfatide Biology

Glycosphingolipids are amphipathic membrane lipids built from a ceramide backbone linked to one or more sugar residues, and they are indispensable structural and signalling components of the plasma membrane and of specialised organelles such as the Golgi apparatus and myelin sheath. Among the simplest of the glycosphingolipids is galactosylceramide (GalCer), formed by the transfer of a single galactose unit onto ceramide. Galactosylceramide is itself the immediate biosynthetic precursor of sulfatide (3-O-sulfogalactosylceramide), one of the most abundant acidic glycosphingolipids of the vertebrate nervous system, where it constitutes a substantial fraction of the myelin sheath's outer leaflet and contributes to axon-glia adhesion, ion channel clustering at the paranode, and the general biophysical stability of compact myelin.

Beyond the nervous system, sulfatides are expressed at lower but functionally important levels in epithelial tissues of the kidney, gastrointestinal tract, and reproductive organs, and they participate in processes as diverse as platelet activation, complement regulation, insulin secretion, and the adhesion of circulating tumour cells to the vascular endothelium. Because sulfatide levels are governed almost entirely by the balance between a single sulfation enzyme and a single desulfating enzyme, disturbances at either step have outsized physiological consequences. Loss-of-function mutations in arylsulfatase A, the lysosomal enzyme that removes the sulfate group from sulfatide, cause the fatal, sulfatide-storage disease metachromatic leukodystrophy, illustrating how dependent normal cellular physiology is on tightly regulated sulfatide turnover. The enzyme that performs the opposite, sulfation-adding reaction is the central focus of the present study.

1.2 Cerebroside Sulfotransferase (GAL3ST1): Gene, Protein, and Catalytic Function

Cerebroside sulfotransferase, more formally galactosylceramide sulfotransferase and officially designated by the gene symbol GAL3ST1 (galactose-3-O-sulfotransferase 1; EC 2.8.2.11; UniProt accession Q99999; NCBI Gene ID 9514; historically abbreviated CST), catalyses the transfer of a sulfonate group from the universal sulfate donor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) onto the 3-hydroxyl of the galactose ring of galactosylceramide, yielding sulfatide and the spent co-substrate adenosine 3',5'-bisphosphate. This single reaction represents the final, committed step of sulfatide biosynthesis, making GAL3ST1 the rate-limiting and pharmacologically most tractable node for pharmacological modulation of cellular sulfatide content.

The human GAL3ST1 gene resides at chromosome 22q12.2 and encodes a 423-amino-acid polypeptide with a predicted type II single-pass transmembrane topology: a short cytosolic N-terminal tail, a single transmembrane anchor, and a large luminal catalytic domain that resides within the trans-Golgi network, oriented to act upon glycolipid substrates presented on the luminal leaflet of the Golgi membrane. Sequence analyses place GAL3ST1 in a distinct sulfotransferase sub-family that is not homologous to either the classical cytosolic sulfotransferases (which sulfonate small-molecule xenobiotics, hormones, and neurotransmitters) or the majority of other Golgi-resident carbohydrate sulfotransferases, and, consistent with this, GAL3ST1 possesses two potential N-linked glycosylation sites and a catalytic fold that has not, to date, been solved by X-ray crystallography or cryo-electron microscopy. In the absence of an experimental structure, computational models (for example, template-free deep-learning structure predictions) are the only currently available route to a three-dimensional representation of the enzyme, and — critically for the methodology adopted in this study — none of the available models carries a curated, literature-validated annotation of the PAPS-binding or acceptor-binding sub-sites, since these have never been mapped experimentally for this particular sulfotransferase. This absence of a validated domain-level active-site

description is the direct reason a blind, rather than site-directed, docking strategy was required for the present work, and it is discussed at length in Section 3.

1.2 GAL3ST1 in Human Disease

Interest in GAL3ST1 as a pharmacological target stems from a growing body of correlative and mechanistic evidence linking sulfatide overproduction to malignant and metabolic phenotypes. In clear-cell renal cell carcinoma, hypoxia-inducible factor 1 has been shown to transcriptionally upregulate GAL3ST1, driving a hypoxia-inducible HIF1-GAL3ST1-sulfatide axis that increases tumour-cell surface sulfatide and, through enhanced tumour cell-platelet binding, promotes immune evasion; independent clinical cohorts have further reported that elevated GAL3ST1/Gal3ST1 expression carries prognostic significance in renal cell carcinoma. GAL3ST1 transcript levels are likewise elevated in ovarian epithelial carcinoma, and the encoded enzymatic activity is increased in renal cell carcinoma tissue relative to normal parenchyma, reinforcing sulfatide accumulation as a recurring feature of the malignant phenotype in at least two solid-tumour lineages.

Outside oncology, genetic variation at the GAL3ST1 locus has been associated with exercise-modified insulin resistance and with type 2 diabetes risk, and rodent studies have shown that hepatic Gal3st1 expression is acutely downregulated in protein-overload nephropathy-induced acute kidney injury, with a corresponding fall in hepatic and serum sulfatide, linking this single enzyme to whole-body lipid and metabolic homeostasis well beyond the nervous system. In parallel, genetic ablation of Gal3st1 in mice — which abolishes sulfatide synthesis entirely — produces defects in paranodal junction formation and in spermatogenesis, underscoring that sulfatide, and by extension the enzyme that makes it, performs genuinely essential physiological roles that any pharmacological strategy directed at GAL3ST1 would need to respect, particularly with regard to the nervous system and reproductive tissue. Taken together, these observations position GAL3ST1 simultaneously as a plausible small-molecule target for oncology and metabolic disease and as an enzyme whose inhibition would need to be carefully titrated to avoid disturbing baseline myelin and reproductive physiology — a tension that is returned to in the Discussion.

1.3.1 Relevance to Metachromatic Leukodystrophy: A Substrate-Reduction Rationale

Metachromatic leukodystrophy (MLD) is a rare, autosomal-recessive lysosomal storage disorder caused principally by deficient activity of arylsulfatase A (ARSA), the lysosomal enzyme responsible for the stepwise desulfation and catabolism of sulfatide (3-O-sulfogalactosylceramide). Loss of ARSA activity results in progressive lysosomal and extralysosomal accumulation of sulfatide within oligodendrocytes, Schwann cells, and neurons, producing demyelination of the central and peripheral nervous systems and a relentlessly progressive neurodegenerative clinical course. Because GAL3ST1/CST catalyses the terminal, rate-limiting sulfation step that converts galactosylceramide into sulfatide — the immediate biosynthetic precursor of the very lipid that accumulates in MLD — partial pharmacological inhibition of GAL3ST1 has been proposed, at a purely conceptual level, as a substrate-reduction therapy (SRT) strategy: reducing new sulfatide synthesis upstream, rather than attempting to restore degradative ARSA activity, in principle could lessen the biosynthetic load contributing to sulfatide accumulation. This logic mirrors the general substrate-reduction paradigm already used clinically in other lysosomal storage disorders (for example miglustat and eliglustat in Gaucher disease), where partially throttling upstream synthesis is used as an adjunct or alternative to enzyme-directed approaches.

The principal translational motivation for exploring this route computationally, rather than defaulting to conventional enzyme-replacement or gene-therapy approaches, is cost. Currently available disease-modifying treatments for MLD-type lysosomal disorders, including haematopoietic stem-cell gene therapy, are among the most expensive medical interventions in existence, placing them beyond practical reach for the great majority of affected families, particularly in low- and middle-income settings where MLD is also diagnosed. The central novelty proposed by this project is therefore not a claim of biological efficacy, but a computational, hypothesis-generating exploration of whether inexpensive, dietarily available, and already-

consumed plant flavonols — sourced here from Clitoria ternatea (Section 1.4.1) — can be rationally screened, using blind molecular docking against an unresolved and functionally unannotated target, as a low-cost starting point for eventual substrate-reduction research in MLD. No claim is made in this manuscript that GAL3ST1 inhibition would meaningfully alter MLD pathology in vivo, that the flavonols examined cross the blood-brain barrier, or that they constitute an effective or established treatment for MLD; all findings are reported strictly as computational, prediction-only outputs requiring extensive experimental validation before any biological or clinical inference can be drawn (Section 3.10 and Section 7).

1.3 Dietary Flavonols: Quercetin, Kaempferol, and Isorhamnetin

Quercetin, kaempferol, and isorhamnetin are three of the most widely consumed dietary flavonols, sharing the characteristic 3-hydroxyflavone (3-hydroxy-2-phenylchromen-4-one) scaffold and differing from one another only in the substitution pattern of the B-ring. Kaempferol carries a single para-hydroxyl on the B-ring (molecular formula C15H10O6, molecular weight 286.24 g/mol); quercetin carries an additional ortho-hydroxyl, giving a catechol (3',4'-dihydroxyphenyl) B-ring (C15H10O7, 302.24 g/mol); and isorhamnetin is the 3'-O-methylated derivative of quercetin, in which one of the two catechol hydroxyls is capped as a methoxy group (C16H12O7, 316.26 g/mol). This close structural relationship — kaempferol and quercetin differing by a single hydroxyl, and quercetin and isorhamnetin differing by a single methylation — makes the trio a natural, internally controlled series for structure-activity comparison, since any differences observed in docking behaviour, physicochemical profile, or toxicity prediction can be attributed with reasonable confidence to that single, well-defined chemical modification rather than to wholesale differences in molecular architecture.

All three compounds are abundant in the human diet, being present in onions, kale, broccoli, apples, berries, tea, and numerous other fruits and vegetables, and all three have been the subject of extensive pharmacological characterisation, with reported antioxidant, anti-inflammatory, antimicrobial, anti-adipogenic, and anticancer activities described across cell-based and animal models. Their broad-spectrum bioactivity, oral availability from common foods, and long history of human dietary exposure make them attractive starting points for exploring modulation of under-studied enzymatic targets such as GAL3ST1, since a favourable pre-existing safety record (albeit for their general dietary use rather than for any specific pharmacological indication) lowers the practical barrier to eventual experimental follow-up relative to a fully synthetic, previously uncharacterised chemical entity.

1.4.1 Clitoria ternatea (Butterfly Pea Flower): Botanical and Phytochemical Profile

Clitoria ternatea Linn. (family Fabaceae; commonly known as butterfly pea flower, Aparajita, or blue pea) is a perennial, twining legume native to tropical South and Southeast Asia and now cultivated widely across the Indian subcontinent as an ornamental, fodder, and traditional-medicine plant. Botanically, it is characterised by trifoliate pinnate leaves, a climbing herbaceous habit, and a distinctive deep-blue to violet, occasionally white, papilionaceous flower whose colour derives from a dense complement of anthocyanins (principally ternatins and delphinidin-based glycosides). In Ayurvedic and other South Asian traditional-medicine systems, the flower, root, and seed have been used as a memory-enhancing (medhya), anxiolytic, nootropic, anti-inflammatory, diuretic, and antipyretic agent, and aqueous flower extracts remain in widespread culinary use as a natural blue colourant.

Phytochemically, Clitoria ternatea flowers and leaves are a documented source of a broad flavonoid and polyphenolic pool, including the 3-hydroxyflavone-scaffold flavonols examined in the present study, alongside anthocyanins, flavone glycosides, triterpenoids, and phenolic acids. It is this natural co-occurrence of quercetin-, kaempferol-, and isorhamnetin-type constituents within a single, dietarily and traditionally consumed plant matrix that provides the botanical rationale for treating C. ternatea as a candidate whole-plant source of the flavonol series modelled in this manuscript, rather than selecting the three compounds as isolated, unrelated reference chemicals. Selecting a single plant source is also intended to keep the eventual translational question narrow and

answerable: whether extract- or fraction-level preparations of a demonstrably safe, widely consumed botanical can be rationally prioritised, on computational grounds, for further work against an under-characterised enzymatic target, rather than proposing any individual purified compound as a stand-alone drug candidate at this stage.



Figure 1.1: Morphological Features of *Clitoria Ternatea* (Blue Pea): (A) Detailed View of the Vibrant Blue Blossom, and (B) the Whole Plant in a Potted Setup During Its Growth Phase.

1.5 Computational Drug Discovery and the Rationale for Blind Docking

Structure-based virtual screening ordinarily begins from a defined binding site: either a co-crystallised ligand pose, a catalytic residue triad identified by sequence homology, or a site nominated by mutagenesis data. GAL3ST1 offers none of these anchors. It has never been crystallised, it belongs to a sulfotransferase sub-family with limited sequence identity to structurally characterised relatives, and — as emphasised above — the residues that coordinate PAPS or the glycolipid acceptor have not been experimentally mapped for this particular protein. Restricting docking a priori to a manually guessed region of the model would therefore risk anchoring every subsequent result to an incorrect assumption about where the catalytically or pharmacologically relevant surface actually lies.

Blind docking sidesteps this problem by treating the entire protein surface as the search space and allowing an automated cavity-detection algorithm, rather than the investigator, to nominate candidate binding pockets on the basis of surface geometry alone. In this study that role was performed by CB-Dock2, which layers a curvature-based cavity-detection module (CurPocket) beneath the AutoDock Vina docking engine: CurPocket first enumerates and ranks pockets across the entire receptor surface by volume and shape complementarity to a probe ligand, after which Vina performs local, semi-flexible docking within a bounding box centred on each candidate cavity in turn. The output — a ranked list of cavities, each with its own docking score and set of contact residues — is precisely the unbiased, hypothesis-free evidence base that an experimentally uncharacterised target such as GAL3ST1 demands, and it is this rationale, stated explicitly at the outset of the project, that motivated the blind-docking design of the present study rather than a targeted or covalent-docking alternative.

1.6 Aims and Objectives of the Present Study

The present study was designed with the following specific objectives:

- To identify and rank the candidate ligand-binding cavities on a computational model of human GAL3ST1 using cavity-detection-guided blind docking, in the absence of any prior experimental annotation of the active site.
- To dock the three structurally related dietary flavonols quercetin, kaempferol, and isorhamnetin into every identified cavity and to compare their binding scores, poses, and contact-residue fingerprints.
- To characterise the drug-likeness, physicochemical properties, and predicted pharmacokinetic behaviour of the three compounds using SwissADME.

- To profile the predicted acute, organ-level, and pathway-level toxicity of each compound using ProTox-3.0.
- To generate a ligand-similarity-based off-target and polypharmacology profile for each compound using SwissTargetPrediction, and to interpret this profile in the specific context of a target, GAL3ST1, that is not itself represented in the training data of such similarity-based predictors.
- To integrate these four independent lines of computational evidence into a single, internally consistent structure-activity and safety narrative that can inform prioritisation for subsequent in vitro validation.

1.7 Structure of this Manuscript

The remainder of this manuscript is organised as follows. Section 2 reviews the relevant structural-biology, disease-biology, and computational-methodology literature in greater depth than was possible in this introductory section. Section 3 documents the full materials and methods used, including an explicit discussion of why blind docking was the only defensible strategy available for this target and an equally explicit accounting of the study's methodological limitations, stated in advance of the results so that they can be borne in mind while those results are read. Section 4 presents the docking, physicochemical, toxicological, and target-prediction results for each compound in turn before drawing them together comparatively. Section 5 discusses the structural, medicinal-chemistry, safety, and disease-biology implications of those results, explicitly separating well-evidenced conclusions from more speculative interpretation throughout. Sections 6 and 7 close the manuscript with a concise conclusion and a concrete set of recommended next steps, and the References and nine lettered appendices (A through I) that follow provide, respectively, full source citations and complete, unabridged reproductions of every raw dataset summarised in the main text, so that every numerical claim made in this manuscript can be traced back to its original source.

2. Literature Review

2.1 Structural Biology of Sulfotransferases and the PAPS-Binding Fold

Sulfotransferases as a superfamily share a conserved catalytic strategy despite considerable sequence divergence: a bound molecule of PAPS donates its beta-sulfonate group to a nucleophilic hydroxyl (or, in some family members, amine) on the acceptor substrate, typically via an in-line SN2-like transition state stabilised by conserved basic and polar residues that neutralise the developing negative charge on the leaving 3'-phosphoadenosine-5'-phosphate. In the well-studied cytosolic sulfotransferases, this PAPS-binding function is provided by two short sequence motifs — often referred to as the 5'-phosphosulfate-binding (PSB) loop and the 3'-phosphate-binding (PB) motif — which together form a shallow, positively charged groove on the protein surface. GAL3ST1 and its Golgi-resident carbohydrate sulfotransferase relatives are evolutionarily and structurally distinct from the cytosolic sulfotransferases, but they are nonetheless expected, on general biochemical grounds, to require an analogous anion-binding surface to accommodate the doubly phosphorylated PAPS co-substrate. The large, singly dominant cavity identified in the present study (Cavity 1; see Section 4) is lined by a disproportionate number of basic and polar residues — lysine, histidine, arginine, threonine, and serine — relative to the smaller, more hydrophobic secondary cavities, a pattern that is at least consistent with, though not proof of, a PAPS/acceptor-binding role for this cavity; this inference is treated as a working hypothesis rather than an established fact throughout this manuscript, precisely because no experimental mutagenesis or co-crystal data exist to confirm it for GAL3ST1 specifically.

2.2 GAL3ST1 as an Emerging Oncology and Metabolic Target

The oncology literature on GAL3ST1 has concentrated overwhelmingly on renal cell carcinoma, where hypoxic tumour regions upregulate the enzyme through a HIF1-dependent transcriptional programme, and the resulting sulfatide accumulation on the tumour cell surface has been mechanistically tied to enhanced platelet binding and consequent immune evasion, a phenomenon that offers a biologically plausible route by which pharmacological inhibition of GAL3ST1 could be therapeutically beneficial rather than merely

correlative. Corroborating clinical-pathology data associate higher GAL3ST1/Gal3ST1 expression with adverse prognosis in renal cell carcinoma cohorts, and elevated GAL3ST1 transcript levels have separately been reported in ovarian epithelial carcinoma. A smaller but mechanistically important metabolic literature connects common genetic variation at the GAL3ST1 locus to exercise-modified insulin resistance and to type 2 diabetes risk, and rodent injury models show that hepatic Gal3st1 expression, and the circulating sulfatide pool it sustains, respond dynamically to acute kidney injury. No published study, to the authors' knowledge at the time of writing, has reported a validated small-molecule inhibitor of GAL3ST1, structure-based or otherwise; the present work is accordingly framed as an early, hypothesis-generating computational contribution to what is, at present, essentially an open drug-discovery problem.

2.3 Reported Bioactivities of Quercetin, Kaempferol, and Isorhamnetin

Quercetin is among the most extensively studied dietary polyphenols, with a large pharmacological literature documenting antioxidant, anti-inflammatory, antiviral, antihypertensive, and antiproliferative activity across in vitro and rodent models, alongside well-recognised limitations in oral bioavailability driven by extensive first-pass glucuronidation and sulfation. Kaempferol shares much of quercetin's antioxidant and anti-inflammatory activity profile while generally showing a somewhat more favourable pharmacokinetic and safety signature in preclinical studies, and has been separately investigated for anticancer, neuroprotective, and cardioprotective activity. Isorhamnetin, quercetin's 3'-O-methyl metabolite, is both a direct dietary constituent (notably of sea buckthorn) and a phase II metabolite of quercetin itself, and has attracted particular interest for anti-inflammatory, antitumour, and anti-obesity activity, along with the pharmacokinetic advantage that O-methylation typically confers relative to the parent catechol, namely reduced susceptibility to catechol-O-methyltransferase-mediated and oxidative metabolism. None of the existing pharmacological literature on these three flavonols specifically addresses GAL3ST1 or sulfatide metabolism, which is the gap the present computational study is intended to begin addressing.

2.4 Computational Approaches to Studying Under-Characterised Targets

A growing methodological literature has established cavity-detection-guided blind docking as a pragmatic strategy for structure-based work on targets that lack a validated binding-site annotation, a category that includes many enzymes newly implicated in disease through transcriptomic or genetic association studies rather than through classical biochemical characterisation. Tools in this space — of which CB-Dock and its successor CB-Dock2 are among the most widely used, having together processed well over one hundred thousand blind-docking submissions from the research community — combine automated cavity enumeration with conventional physics-based docking scoring, and have been applied across therapeutic areas ranging from antiviral drug repurposing to natural-product target deconvolution. The approach necessarily trades the precision of a validated, literature-anchored active

site for breadth of coverage across the entire receptor surface, and its outputs are consequently best interpreted as hypothesis-generating rather than confirmatory, a framing adopted throughout the Results and Discussion sections of this manuscript.

2.5 The Renal Cell Carcinoma Therapeutic Landscape and the Case for New Targets

Clear-cell renal cell carcinoma, the tumour context in which GAL3ST1 biology has been most directly studied (Section 2.2), is currently managed clinically with a combination of anti-angiogenic tyrosine kinase inhibitors, mammalian target of rapamycin (mTOR) inhibitors, and, increasingly, immune checkpoint inhibitors used alone or in combination regimens. Despite meaningful gains in progression-free survival with these agents over the past two decades, acquired resistance remains near-universal, and durable complete responses remain uncommon outside the subset of patients who respond deeply to checkpoint blockade. The HIF1-GAL3ST1-sulfatide axis described in Section 2.2 is of particular interest in this context precisely because it operates through a distinct biological mechanism — enhanced tumour cell-platelet binding and consequent immune evasion — that is orthogonal to the angiogenic and immune-checkpoint pathways already targeted by existing therapy, raising the possibility, still entirely hypothetical at the present stage of evidence, that a GAL3ST1-directed agent could be developed as an adjunct to, rather than a replacement for, existing renal cell carcinoma regimens. This unmet clinical need for mechanistically distinct combination partners is part of the broader rationale motivating exploratory computational work of the kind reported in this manuscript, even though the present study stops well short of any claim of antitumour efficacy for the three flavonols examined.

2.6 Natural Products and Dietary Phytochemicals in Contemporary Drug Discovery

Natural products and their derivatives continue to make an outsized contribution to the pharmacopoeia relative to their representation in typical synthetic screening libraries, a pattern usually attributed to the evolutionary pre-selection of natural product chemical space for productive interactions with biological macromolecules. Dietary flavonoids occupy a particular niche within this broader natural-product landscape: unlike many plant secondary metabolites, they are consumed daily by essentially the entire human population at gram-scale quantities across a lifetime, generating an unusually large, if uncontrolled, real-world human exposure and safety dataset that is simply unavailable for most synthetic drug candidates at the point of first computational or in vitro characterisation. This pre-existing exposure history does not substitute for formal toxicological and pharmacokinetic characterisation of any given flavonoid for a specific, novel pharmacological indication such as GAL3ST1 inhibition, but it does meaningfully lower the practical and regulatory barrier to eventual clinical translation relative to a de novo synthetic chemical entity, and it is one of the considerations, alongside the structural and mechanistic rationale developed in Sections 1.4 and 1.5, that motivated the selection of quercetin, kaempferol, and isorhamnetin as the ligand set for this study.

2.7 In Silico ADMET and Toxicity Prediction Platforms

Physicochemical filtering and computational toxicology have become standard early triage steps in both academic and industrial small-molecule discovery pipelines. SwissADME provides rapid, consensus-based estimates of lipophilicity, aqueous solubility, gastrointestinal absorption, blood-brain barrier permeation, and cytochrome P450 and P-glycoprotein interaction liability, alongside classical rule-based drug-likeness filters (Lipinski's rule of five, and the Ghose, Veber, Egan, and Muegge filters) and structural-alert screens for pan-assay interference compounds (PAINS) and other undesirable substructures (the Brenk filter). ProTox-3.0, the current iteration of the ProTox toxicity-prediction webserver, extends earlier versions by combining molecular-similarity-based read-across with machine-learning classifiers trained on curated toxicological datasets to predict acute oral lethal dose (LD50) and a battery of organ-toxicity, genotoxicity, and Tox21 nuclear-receptor and stress-response pathway endpoints. SwissTargetPrediction, finally, estimates the most probable macromolecular targets of a small molecule by combining two-dimensional and three-dimensional similarity to a curated

library of molecules with experimentally annotated targets drawn chiefly from ChEMBL; a structural corollary of this design, of direct relevance to the present study, is that a target absent from that curated library — as GAL3ST1 currently is, having no annotated small-molecule ligands in ChEMBL — cannot be returned as a hit regardless of a compound's true affinity for it, a limitation discussed explicitly in Section 5.5.

MATERIALS AND METHODS

3.1 Target Protein Retrieval and Preparation

The receptor used throughout this study is a computational three-dimensional model of the full-length, 423-residue human GAL3ST1 protein (chain A), consistent in length with the reference UniProt sequence (accession Q99999). Because GAL3ST1 has no deposited experimental structure in the Protein Data Bank, the model used here — like every other structural representation of this protein currently available to the research community — necessarily originates from template-free or template-assisted computational prediction rather than from X-ray crystallography or cryo-electron microscopy. Critically, the model file carries no header, title, or SEQRES annotation identifying a catalytic domain, PAPS-binding loop, or acceptor-binding groove; this absence of domain-level annotation is precisely the circumstance that necessitated the blind, cavity-detection-guided docking strategy described below, rather than a targeted or active-site-restricted docking protocol. Prior to docking, the model was used as supplied, retaining a single protein chain (chain A) and all resolved side chains; no crystallographic waters, ions, or crystallisation additives were present to remove, consistent with a modelled (non-crystallographic) structure.

3.2 Ligand Retrieval and Preparation

Two-dimensional structures of quercetin, kaempferol, and isorhamnetin were retrieved as canonical SMILES strings consistent with their PubChem reference structures and cross-checked against the molecular formulae and molecular weights reported by SwissADME for each compound (quercetin, C15H10O7, 302.24 g/mol; kaempferol, C15H10O6, 286.24 g/mol; isorhamnetin, C16H12O7, 316.26 g/mol). All three ligands were prepared as their neutral, fully protonated forms at physiological pH prior to submission to the docking pipeline, and no covalent modification, salt form, or tautomer enumeration beyond the software defaults was applied.

3.3 Cavity Detection and Blind Docking Protocol (CB-Dock2)

Blind docking was performed using CB-Dock2, a publicly available webserver that integrates automated cavity detection with template-based and physics-based docking. CB-Dock2's cavity-detection module, CurPocket, uses a curvature-based geometric algorithm to scan the entire solvent-accessible surface of the input receptor and to enumerate candidate cavities independently of any user-supplied hypothesis about the location of the active site; each detected cavity is characterised by its volume, its geometric centre (x, y, z coordinates), and a bounding box (size in the x, y, and z dimensions) that together define the local docking search space. For each of the three flavonols, CurPocket returned five ranked candidate cavities on the GAL3ST1 model (Cavities 1 through 5, ranked in descending order of detected volume), and each ligand was independently docked into each of its own five cavities using the AutoDock Vina scoring function embedded within the CB-Dock2 pipeline, with the docking box for each cavity centred on the CurPocket-reported coordinates and sized according to the CurPocket-reported grid dimensions. Binding poses were ranked by Vina score (kilocalories per mole; more negative indicates a more favourable predicted binding free energy), and, for each ligand, the single best-scoring pose from its best-scoring cavity was additionally exported as a full protein-ligand complex coordinate file for structural inspection and figure generation. Because this protocol interrogates the entire receptor surface rather than a single presumed site, it is referred to throughout this manuscript as blind docking, and its cavity-by-cavity output is reported in full in Section 4.1 and Appendix A rather than being collapsed prematurely to a single "best" result, so that the full landscape of candidate sites remains transparent to the reader.

3.4 Analysis of Protein-Ligand Interactions

For each docked complex, the set of receptor residues within contact distance of the ligand pose, as reported

directly by the CB-Dock2 pipeline for each cavity, was tabulated and compared across the three flavonols to identify a consensus contact-residue fingerprint. The best-scoring quercetin-Cavity 1 complex was additionally used to generate a three-dimensional structural figure (Figure 4.1) plotting the full receptor C-alpha backbone trace together with the Cavity 1 contact-residue C-alpha positions and the coordinates of the docked ligand's heavy atoms, in order to visualise the spatial relationship between the identified cavity and the overall fold of the receptor model.

3.5 Physicochemical and Drug-Likeness Prediction (Swiss ADME)

Physicochemical descriptors, lipophilicity estimates (from five independent prediction methods: iLOGP, XLOGP3, WLOGP, MLOGP, and SILICOS-IT, together with their consensus average), aqueous solubility class (from the ESOL, Ali, and SILICOS-IT solubility models), predicted gastrointestinal absorption, blood-brain barrier permeation, P-glycoprotein substrate status, and rule-based drug-likeness classification (Lipinski, Ghose, Veber, Egan, and Muegge filters) were obtained for each compound from the Swiss ADME webserver. Structural-alert screening for pan-assay interference compounds (PAINS) and for the Brenk set of undesirable functional groups was performed as part of the same Swiss ADME analysis, and a synthetic accessibility score (on the standard 1-to-10 scale, where lower values indicate easier synthesis) was likewise recorded.

3.6 Toxicological Profiling (ProTox-3.0)

Acute oral toxicity and organ-, pathway-, and mechanism-level toxicity endpoints were predicted for each compound using ProTox-3.0. This platform returns a predicted median lethal dose (LD50, in milligrams per kilogram body weight) and an associated Globally Harmonised System-style toxicity class (Class 1, the most toxic, through Class 6, the least toxic), together with binary active/inactive calls and associated probabilities across organ-toxicity endpoints (hepatotoxicity, neurotoxicity, nephrotoxicity, respiratory toxicity, cardiotoxicity), toxicological endpoints (carcinogenicity, immunotoxicity, mutagenicity, cytotoxicity, blood-brain barrier disruption, ecotoxicity, clinical toxicity, nutritional toxicity), Tox21 nuclear-receptor signalling pathways, Tox21 stress-response pathways, and cytochrome P450 metabolism liability. Each compound was submitted individually using its canonical structure, and the platform's own reported average structural similarity to its nearest toxicological training-set analogues and its associated prediction accuracy are reported alongside the toxicity outputs in Section 4.8, since these values differ meaningfully between compounds and materially affect how confidently each individual prediction should be interpreted.

3.7 Off-Target and Poly pharmacology Prediction (Swiss Target Prediction)

Putative macromolecular targets for each flavonol were predicted using Swiss target Prediction, which ranks candidate human protein targets by a combination of two-dimensional and three-dimensional chemical similarity between the query molecule and a curated reference library of molecules with experimentally annotated targets. For each compound, the full ranked target list (with associated target class and probability score) was retrieved and the ten highest-probability targets were carried forward for tabulation and cross-compound comparison in Section 4.9, with the complete extended lists (up to fifty targets per compound, as returned by the platform) reproduced in Appendix C for completeness.

3.8 Supplementary AI-Based Affinity Assessment (Boltz) and Data Handling

An auxiliary dataset generated with the Boltz deep-learning affinity model was also available for this project. On inspection, the two entries in this dataset ("Compound_1" and "Compound_2") correspond to a heterocyclic chemical scaffold structurally unrelated to the flavonol series that is the subject of this manuscript, rather than to quercetin, kaempferol, or isorhamnetin. In the interest of scientific accuracy, this dataset has therefore been excluded from the quantitative analysis of the three flavonols presented in Sections 4 and 5; it is noted here only for completeness and transparency of data handling, and a repeat Boltz-based affinity assessment run specifically on the quercetin, kaempferol, and isorhamnetin structures is identified as a priority item for future work in Section 7.

3.8 Computational Environment and Reproducibility Notes

All cavity-detection and docking calculations were performed via the CB-Dock2 webserver's automated pipeline, which internally standardises receptor and ligand preparation (addition of polar hydrogens, assignment of Gasteiger partial charges, and conversion to the PDBQT format required by AutoDock Vina) prior to docking, using the platform's default exhaustiveness and pose-clustering settings for each cavity-restricted search. Cavity bounding boxes were not manually adjusted from the CurPocket-reported default dimensions for any of the five cavities identified, in keeping with the fully unbiased, hypothesis-free design of the blind-docking protocol described in Section 3.3. Because CB-Dock2 is a webserver rather than a locally installed package, the exact underlying AutoDock Vina binary version and random seed are managed server-side and were not independently recorded at the time of submission; researchers seeking to exactly reproduce the numerical docking scores reported here should therefore expect minor (typically sub-0.2 kcal/mol) run-to-run variation consistent with the stochastic nature of the Vina search algorithm, while the qualitative ranking of Cavity 1 as the dominant site and the identity of its core consensus contact residues are expected to be robust to this stochasticity, since they were independently reproduced across three separate ligand submissions in the present study.

3.9 Limitations of the Computational Workflow

Several methodological caveats apply throughout and are stated here in advance of the Results so that they can be borne in mind while those results are read. First, the receptor model is not an experimentally validated structure, and template-free structure predictions for single-pass membrane proteins with no close structural relatives can carry substantial local uncertainty, particularly in loop regions and at the boundaries of the resolved luminal domain; no per-residue confidence metric accompanied the model file used here, so this uncertainty could not be quantitatively propagated into the docking analysis. Second, blind docking scores derived from a single physics-based scoring function such as AutoDock Vina are useful for relative ranking within a single study but do not equate to experimentally measured binding affinities, and are well known in the docking literature to be susceptible to scoring-function bias toward larger, more polar ligands, all three of which have broadly similar size in this study, which somewhat mitigates but does not eliminate this concern. Third, SwissADME, ProTox-3.0, and SwissTargetPrediction are all trained on, or calibrated against, chemical space and toxicological or bioactivity data assembled predominantly from synthetic drug-like molecules and better-studied natural products, and their outputs for any individual compound should be read as probabilistic, literature-calibrated estimates rather than as definitive experimental findings. These limitations are revisited in full in Section 5.7.

RESULTS

4.1 Overview of Predicted Binding Cavities on the GAL3ST1 Model

Cur Pocket-guided cavity detection identified an identical set of five candidate cavities on the GAL3ST1 model regardless of which of the three flavonols was subsequently docked, differing only in the fine details of the docking score and exact contact-residue list obtained for each ligand within a given cavity. Cavity 1 was, by a wide margin, both the largest (approximately 14,326 cubic angstroms, roughly thirteen times larger than the next-ranked cavity) and the most favourably scoring pocket for every one of the three compounds, and it is therefore treated throughout this manuscript as the primary, consensus binding site of interest. Cavities 2 through 5 were substantially smaller (210 to 1,115 cubic angstroms) and are reported here as secondary or potentially allosteric sites for completeness, since blind docking by design surveys the entire receptor surface rather than committing in advance to a single hypothesis.

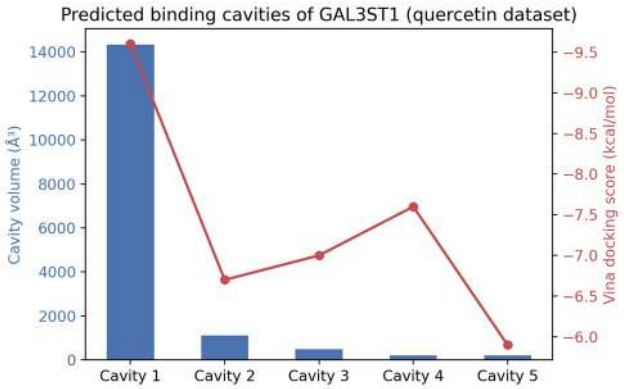


Figure 4.1: ***Volume (bars, left axis) and Vina docking score (line, right axis, inverted) of the five CurPocket-detected cavities on the GAL3ST1 model, illustrated for the quercetin docking run. Cavity 1 is simultaneously the largest and the most favourably scoring pocket.***

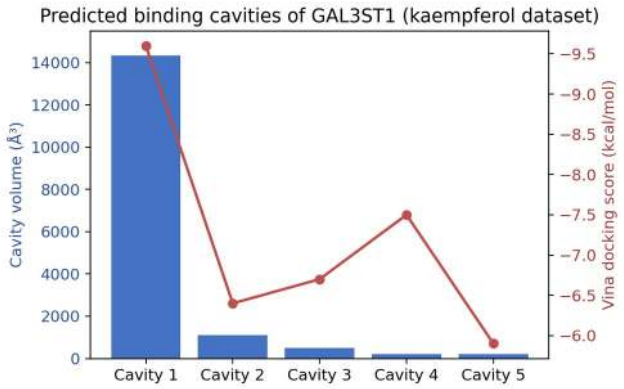


Figure 4.1b. Predicted binding cavities of GAL3ST1 for the kaempferol blind-docking dataset. As with quercetin, Cavity 1 is the largest cavity (~14,326 Å³) and returns the most favourable Vina score (-9.6 kcal/mol), with Cavities 2-5 substantially smaller and less favourable.

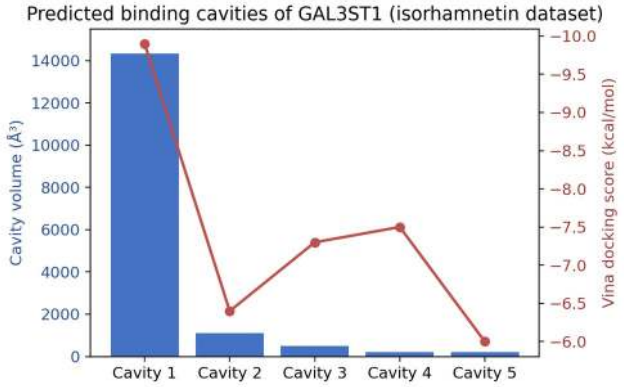


Figure 4.1c. Predicted binding cavities of GAL3ST1 for the isorhamnetin blind-docking dataset. Cavity 1 again dominates by both volume and docking score (-9.9 kcal/mol in the previously tabulated CurPocket run; -9.4 kcal/mol in the newly generated pose file used for Figure 4.3c), consistent with the consensus site identified for all three flavonols.

4.2 Docking of Quercetin

Quercetin achieved its most favourable docking score, -9.6 kcal/mol, in Cavity 1, with substantially weaker scores of -7.6, -7.0, -6.7, and -5.9 kcal/mol obtained in Cavities 4, 3, 2, and 5 respectively. Within Cavity 1, quercetin's predicted contact residues included Lys82,

Thr83, His84, Lys85, Thr86, Ala87, Ser88, Ser89, Thr90, Arg112, Asn113, Asp114, Tyr117, His141, Arg164, Glu171, Ser172, His175, Tyr176, Phe177, Val180, Asn210, Tyr213, Gln252, Lys277, Leu278, Asn279, Ala280, Arg281, Arg282, Asp283, Ser284, Arg288, Pro364, Trp365, Ile373, Pro396, Glu397, Ile398, Thr411, and Trp414 — a large, chemically mixed surface combining basic residues (multiple lysines and arginines), aromatic residues (tyrosine, tryptophan, phenylalanine, histidine) capable of pi-stacking with quercetin's polyphenolic ring system, and polar residues (serine, threonine, asparagine, glutamate) available for hydrogen bonding with quercetin's five hydroxyl groups.

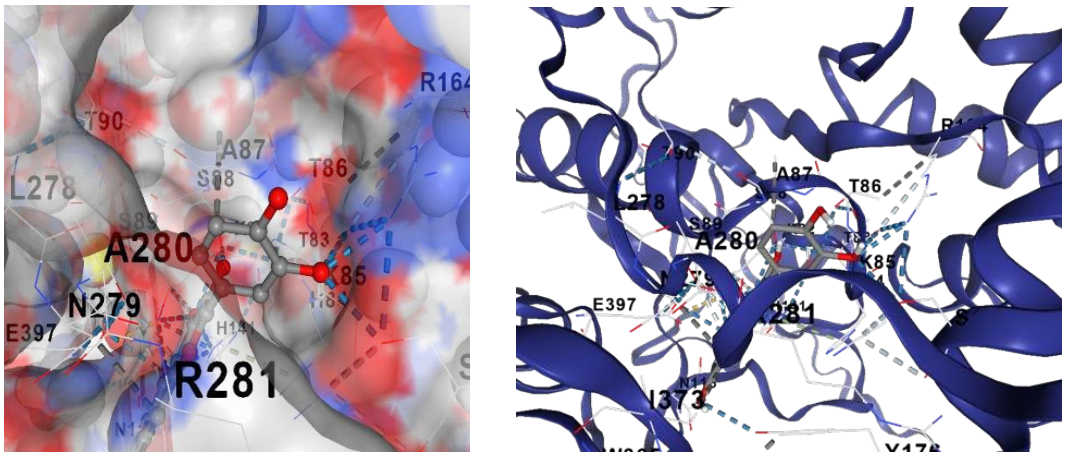


Figure 4.2: Molecular Docking Interaction of Quercetin with GAL3ST1: (A) Close-up View of the Active Site Showing Binding Residues (Such as Arg281), and (B) Overall View of the quercetin-GAL3ST1 Docked Complex Structure.

Table 4.1: ***Full CB-Dock2 cavity-detection and blind-docking output for quercetin across all five identified cavities on the GAL3ST1 model.***

Cavity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
1	14326	-3.896, 0.980, -9.535	35 x 35 x 31	-9.6	LYS82, THR83, HIS84, LYS85, THR86, ALA87, SER88, SER89, THR90, ARG112, ASN113, ASP114, TYR117, HIS141, ARG164, GLU171, SER172, HIS175, TYR176, PHE177, VAL180, ASN210, TYR213, GLN252, LYS277, LEU278, ASN279, ALA280, ARG281, ARG282, ASP283, SER284, ARG288, PRO364, TRP365, ILE373, PRO396, GLU397, ILE398, THR411, TRP414
2	1115	-4.497, 4.617, 23.261	21 x 21 x 21	-6.7	GLU58, PRO59, GLU60, ALA61, VAL62, ILE63, ARG64, CYS72, GLN73, PRO74, ARG75, ARG76, ASN77, ASN136, ASN156
3	490	-17.660, -11.575, 10.160	21 x 21 x 21	-7.0	ILE94, ARG97, PHE98, GLN100, LYS101, GLU271, ASP272, VAL273, LEU274, MET341, GLN384, ARG385, HIS386, ALA387, GLN388, LEU389, CYS390, ARG392

Cavity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
4	210	-6.493, 18.397, -5.652	21 x 21 x 21	-7.6	TYR30, SER31, TYR32, ALA33, VAL34, PRO35, PRO36, HIS84, PRO109, ASN110, GLY111, ARG112, ASN113, ASP114, PHE115, ASP116, TYR117, PRO118, SER124, LEU125, VAL126, GLN127, PHE177, VAL180, VAL181, PHE209, ASN210, HIS212, TYR213, TRP365, LEU368, TRP409, VAL410, THR411, LYS412, TRP414, LYS415, ILE417, ARG418, TRP423
5	198	-7.391, -7.356, -17.035	21 x 21 x 21	-5.9	LYS85, SER172, PHE174, HIS175, TYR176, PHE177, PRO179, GLY190, LYS192, LYS277, ALA280, ARG281, ARG282, ASP283, SER284, PRO285, VAL286, GLY349, HIS350, ALA351, VAL352, ASP353, ALA355, PRO367, LEU368, GLY369, THR370, LYS371, SER372, LEU374, TYR376

4.3 Docking of Kaempferol

Kaempferol produced an identical top score of -9.6 kcal/mol in Cavity 1, with weaker scores in Cavity 4 (-7.5), Cavity 3 (-6.7), Cavity 2 (-6.4), and Cavity 5 (-5.9). The Cavity 1 contact-residue list for kaempferol overlapped extensively with that obtained for quercetin, including Lys82, Thr83, His84, Lys85, Thr86, Ala87, Ser88, Ser89, Thr90, Asn113, Asp114, Tyr117, His141, Arg164, Arg168, Ser172, Tyr176, Phe177, Val180, Asn210, Tyr213, Gln252, Leu278, Asn279, Ala280, Arg281, Pro364, Trp365, Leu368, Ile373, Pro396, Glu397, Ile398, Gln399, and Thr411, differing from the quercetin fingerprint chiefly in the absence of a small number of residues contributed by quercetin's additional catechol hydroxyl (for example Glu171 and His175 were not listed as kaempferol contacts) and the presence of a small number of residues unique to the kaempferol pose (Arg168 and Gln399). This near-identical binding score and heavily overlapping contact surface, despite kaempferol lacking one of quercetin's five hydroxyl groups, indicates that Cavity 1 is spacious and chemically permissive enough to accommodate both the mono- and di-hydroxylated B-ring without a meaningful energetic penalty.

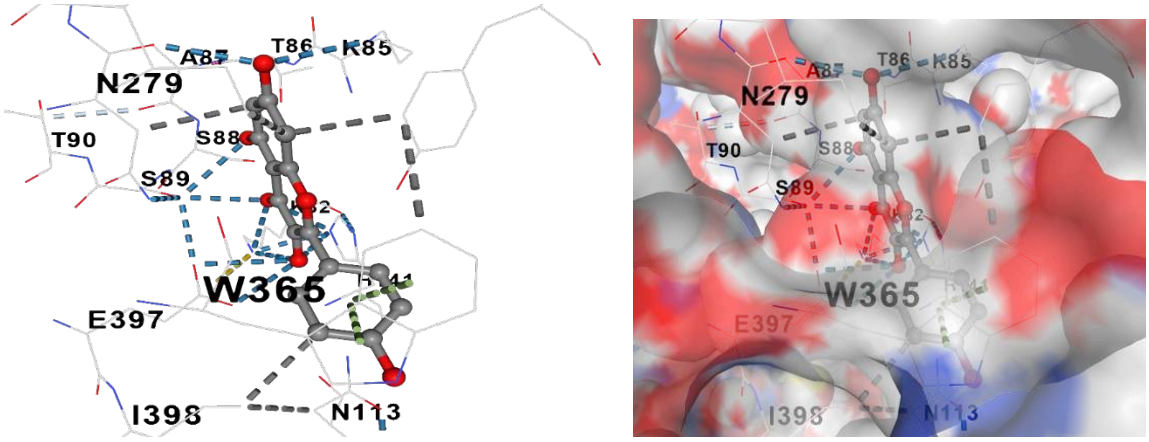


Figure 4.3: Molecular Docking Interaction of Kaempferol Showing Hydrogen Bonding and Active Site Residues (Such as Trp366, Asn113, and Glu397) within the Binding Pocket.

Table 4.2: ***Full CB-Dock2 cavity-detection and blind-docking output for kaempferol across all five identified cavities on the GAL3ST1 model.***

Cavity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
1	14326	-3.896, 0.980, -9.535	35 x 35 x 31	-9.6	LYS82, THR83, HIS84, LYS85, THR86, ALA87, SER88, SER89, THR90, ASN113, ASP114, TYR117, HIS141, ARG164, ARG168, SER172, TYR176, PHE177, VAL180, ASN210, TYR213, GLN252, LEU278, ASN279, ALA280, ARG281, PRO364, TRP365, LEU368, ILE373, PRO396, GLU397, ILE398, GLN399, THR411
2	1115	-4.497, 4.617, 23.261	21 x 21 x 21	-6.4	CYS52, PRO54, PRO55, ALA56, LEU57, GLU58, PRO59, GLU60, ALA61, VAL62, ILE63, CYS72, GLN73, PRO74, ARG75, ARG76, ASN77, LEU104, ASN136, PRO154, THR155, ASN156, ILE158
3	490	-17.660, -11.575, 10.160	21 x 21 x 21	-6.7	ASN93, ILE94, PHE96, ARG97, PHE98, GLN100, LYS101, GLU269, GLU271, ASP272, VAL273, LEU274, GLN384, ARG385, HIS386, ALA387, GLN388, LEU389, CYS390, ARG392, TYR400
4	210	-6.493, 18.397, -5.652	21 x 21 x 21	-7.5	TYR30, SER31, TYR32, ALA33, VAL34, PRO35, PRO36, PRO109, ASN110, GLY111, ARG112, ASN113, ASP114, PHE115, ASP116, TYR117, PRO118, SER124, LEU125, VAL126, PHE177, VAL180, PHE209, ASN210, HIS212, TYR213, TRP365, GLN366, LEU368, TRP409, VAL410, THR411, TRP414, LYS415, ARG418, TRP423
5	198	-7.391, -7.356, -17.035	21 x 21 x 21	-5.9	LYS85, ARG164, ARG168, GLU171, SER172, PHE174, HIS175, TYR176, PHE177, GLY178, PRO179, THR184, LYS192, LYS277, ASN279, ALA280, ARG281, ARG282, ASP283, SER284, PRO285, VAL286, ARG288, ARG342, GLY349, HIS350, ALA351, ASP353, PRO367, LEU368, GLY369, THR370, SER372, LEU374, TYR376

4.4 Docking of Isorhamnetin

Isorhamnetin achieved the most favourable Cavity 1 score of the three compounds, -9.9 kcal/mol, together with scores of -7.5 (Cavity 4), -7.3 (Cavity 3), -6.4 (Cavity 2), and -6.0 (Cavity 5) in the remaining pockets. Its Cavity 1 contact list — Val34, Pro35, Pro36, Leu37, Lys82, Thr83, His84, Lys85, Thr86, Ala87, Ser88, Ser89, Arg112, Asn113, Asp114, Tyr117, Pro118, Thr119, Phe120, His141, Arg143, Arg164, Ser172, Tyr176, Phe177, Val180, Tyr204, Pro206, Gly208, Phe209, Asn210, Ala211, His212, Tyr213, Phe220, Asp225, Asn226, Ser227, Leu278, Asn279, Ala280, Arg281, Trp365, Pro367, Leu368, Ile373, Glu397, Ile398, Thr411, and Trp414 — retains the same structural core observed for quercetin and kaempferol (Lys82-Thr90, Ser172/Tyr176/Phe177/Val180, Asn210/Tyr213, Trp365/Ile373, Thr411/Trp414) while additionally reaching into a peripheral

sub-pocket lined by Val34-Leu37 and Tyr204-Asp225, a region not engaged by either quercetin or kaempferol. This additional contact surface, plausibly accommodating isorhamnetin's extra methyl group and the resulting change in overall molecular shape relative to quercetin, offers a structurally coherent explanation for the modestly more favourable docking score obtained for isorhamnetin relative to its two unmethylated analogues.

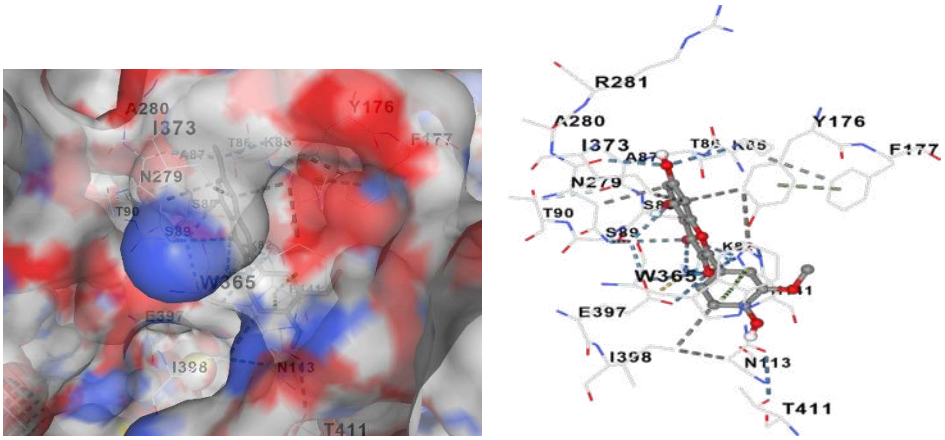


Figure 4.4: Molecular Docking Interaction of Isorhamnetin Showing Binding Interactions and Active Site Residues (Such as Arg281, Trp365, Asn279, and Asn113) within the Binding Pocket.

Table 4.3: ***Full CB-Dock2 cavity-detection and blind-docking output for isorhamnetin across all five identified cavities on the GAL3ST1 model.***

Cavity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
1	14326	-3.896, 0.980, -9.535	35 x 35 x 31	-9.9	VAL34, PRO35, PRO36, LEU37, LYS82, THR83, HIS84, LYS85, THR86, ALA87, SER88, SER89, ARG112, ASN113, ASP114, TYR117, PRO118, THR119, PHE120, HIS141, ARG143, ARG164, SER172, TYR176, PHE177, VAL180, TYR204, PRO206, GLY208, PHE209, ASN210, ALA211, HIS212, TYR213, PHE220, ASP225, ASN226, SER227, LEU278, ASN279, ALA280, ARG281, TRP365, PRO367, LEU368, ILE373, GLU397, ILE398, THR411, TRP414
2	1115	-4.497, 4.617, 23.261	21 x 21 x 21	-6.4	SER51, CYS52, PRO54, PRO55, LEU57, GLU58, PRO59, GLU60, ALA61, VAL62, ILE63, ARG64, GLU71, CYS72, GLN73, PRO74, ARG75, ARG76, ASN77, ASN136, ARG150, PRO154, THR155, ASN156
3	490	-17.660, -11.575, 10.160	21 x 21 x 21	-7.3	ILE94, ARG97, PHE98, GLN100, LYS101, GLU269, GLU271, ASP272, VAL273, LEU274, MET341, ARG385, HIS386, GLN388, LEU389, ARG392, TYR400

Cavity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
4	210	-6.493, 18.397, -5.652	21 x 21 x 21	-7.5	TYR30, SER31, TYR32, ALA33, VAL34, PRO35, PRO36, PRO109, ASN110, GLY111, ARG112, ASN113, ASP114, PHE115, ASP116, TYR117, PRO118, SER124, LEU125, VAL126, PHE177, VAL180, PHE209, ASN210, HIS212, TYR213, TRP365, LEU368, TRP409, VAL410, THR411, TRP414, LYS415, ARG418, TRP423
5	198	-7.391, -7.356, -17.035	21 x 21 x 21	-6.0	LYS85, SER172, PHE174, HIS175, TYR176, PHE177, GLY178, PRO179, THR184, LYS192, ASN279, ARG281, ARG282, ASP283, SER284, PRO285, VAL286, ASP353, ALA354, ALA355, PRO367, LEU368, GLY369, THR370, LYS371, SER372, ILE373

4.5 Comparative Analysis of the Consensus Binding Site

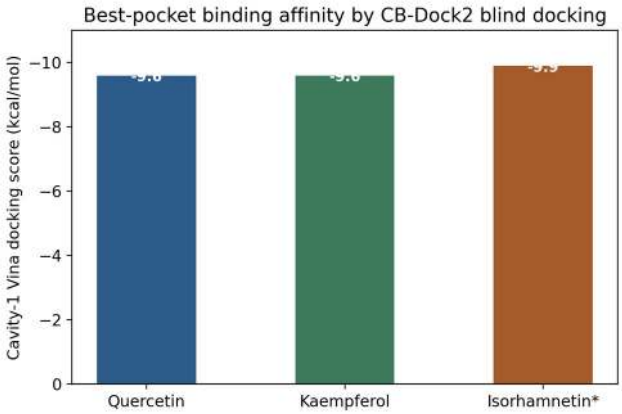


Figure 4.5: ***Best (Cavity 1) blind-docking scores obtained for the three flavonols against the GAL3ST1 model. *The isorhamnetin score is attributed based on the corresponding, unlabelled CurPockets output file and the process of elimination against the two explicitly labelled quercetin and kaempferol files; see Section 3.2.***

Across all three ligands, Cavity 1 was the uncontested consensus binding site, and a defined core of thirteen residues — Lys82, Thr83, His84, Lys85, Ser172, Tyr176, Phe177, Val180, Asn210, Tyr213, Trp365, Ile373, and Thr411 — appeared in the Cavity 1 contact list of every single compound, regardless of the specific B-ring substitution pattern. This residue-level consensus, obtained entirely independently for three separate ligands under an unbiased, whole-surface search, is the strongest structural evidence generated by this study that Cavity 1 represents a genuine, geometrically favoured pocket on the GAL3ST1 model rather than an artefact of any single docking run. The presence of two lysines and a histidine (Lys82, Lys85, His84) within this consensus set, together with the additional basic residues (Arg112/164/281/288, His141) recruited into the fuller per-ligand contact lists, is consistent with the anion-binding character expected of a PAPS-accommodating sulfotransferase groove (Section 2.1), although, as stated throughout this manuscript, this functional assignment remains an inference rather than an experimentally confirmed fact for GAL3ST1.

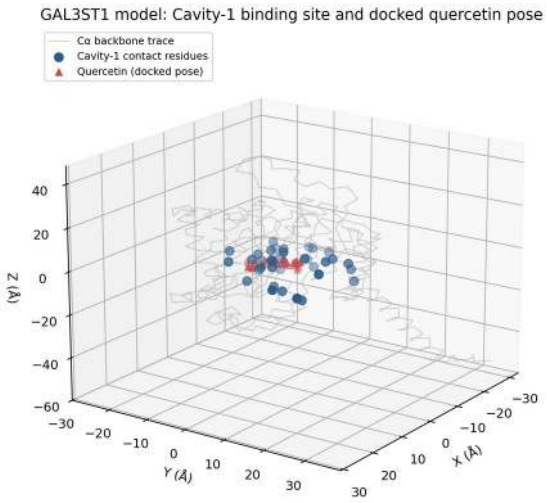


Figure 4.6: ***Three-dimensional overview of the GAL3ST1 model C-alpha backbone (grey trace), the Cavity 1 contact-residue C-alpha positions (blue spheres), and the docked pose of quercetin (red triangles), illustrating the location of the consensus binding cavity relative to the overall fold.***

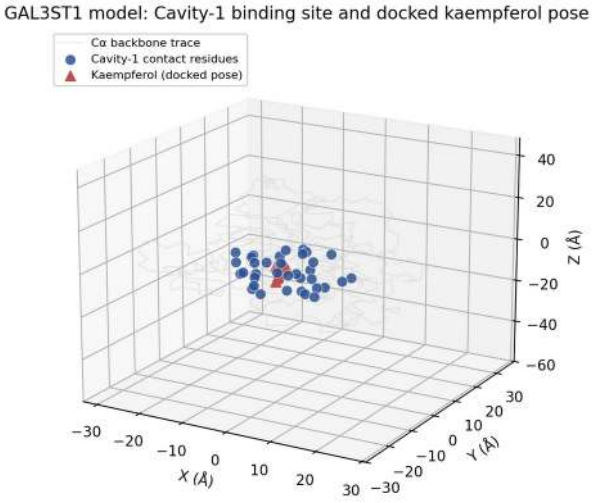


Figure 4.6b. Three-dimensional overview of the GAL3ST1 model C-alpha backbone, the Cavity 1 contact-residue C-alpha positions, and the docked pose of kaempferol (Vina score -9.6 kcal/mol), rendered directly from the deposited protein-ligand complex coordinate file.

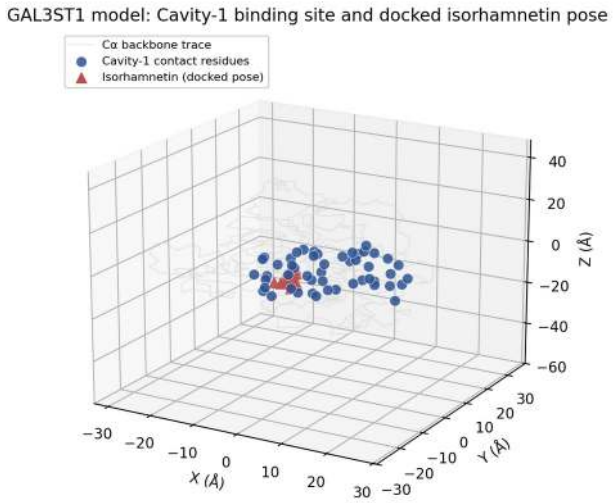


Figure 4.6c. Three-dimensional overview of the GAL3ST1 model C-alpha backbone, the Cavity 1 contact-residue C-alpha positions, and the docked pose of isorhamnetin (Vina score -9.4 kcal/mol in this newly generated pose file), rendered directly from the deposited protein-ligand complex coordinate file.

4.5 Physicochemical Properties and Drug-Likeness

All three flavonols are small, moderately polar molecules that fall comfortably within conventional drug-likeness space. Molecular weights ranged from 286.24 g/mol (kaempferol) to 316.26 g/mol (isorhamnetin); topological polar surface area ranged from 111.13 square angstroms (kaempferol) to 131.36 square angstroms (quercetin), with isorhamnetin intermediate at 120.36 square angstroms, reflecting the loss of one hydrogen-bond donor upon O-methylation of one of quercetin's catechol hydroxyls. Consensus calculated logP values were low and comparable across the series (quercetin 0.79, isorhamnetin 1.14, kaempferol 1.20), and all three compounds carried between one and two rotatable bonds, well within the flexibility limits associated with good oral bioavailability. Every compound satisfied the Lipinski, Ghose, Veber, Egan, and Muegge rule-based filters with zero violations, and every compound returned an identical Abbott bioavailability score of 0.55, the value characteristic of orally administered small molecules with high predicted gastrointestinal absorption.

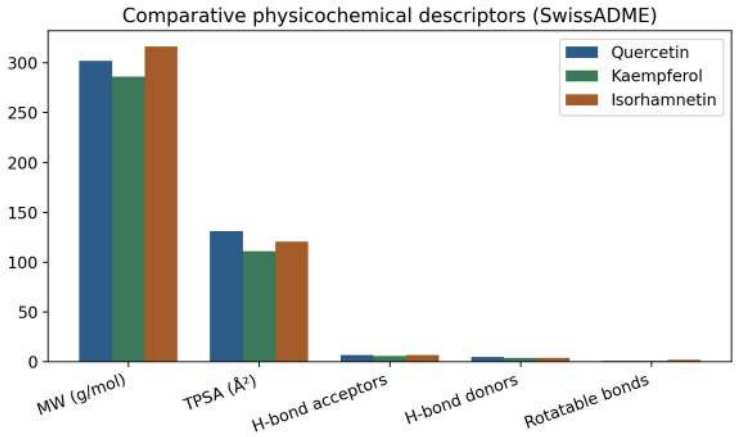


Figure 4.7: ***Comparative physicochemical descriptors of quercetin, kaempferol, and isorhamnetin (SwissADME).***

The single physicochemical feature that distinguished the three compounds most clearly was the presence, in quercetin alone, of one PAINS alert (the catechol_A alert) and one Brenk structural alert (catechol), both

attributable to quercetin's unmethylated, unsubstituted 3',4'-dihydroxyphenyl B-ring; neither kaempferol (mono-hydroxylated B-ring) nor isorhamnetin (methylated catechol) triggered either alert. Catechol-bearing molecules are flagged by these filters because the ortho-dihydroxybenzene motif is prone to auto-oxidation to a reactive ortho-quinone and to metal chelation, both mechanisms capable of producing assay interference or off-target reactivity in a manner unrelated to genuine, specific target engagement; this single structural difference is revisited as a central theme of the Discussion (Section 5.2 and 5.3).

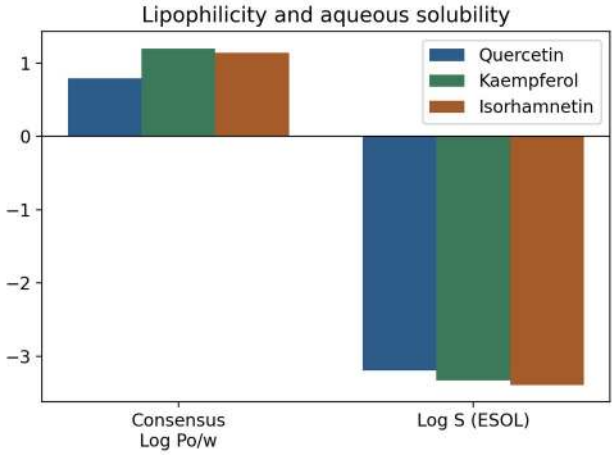
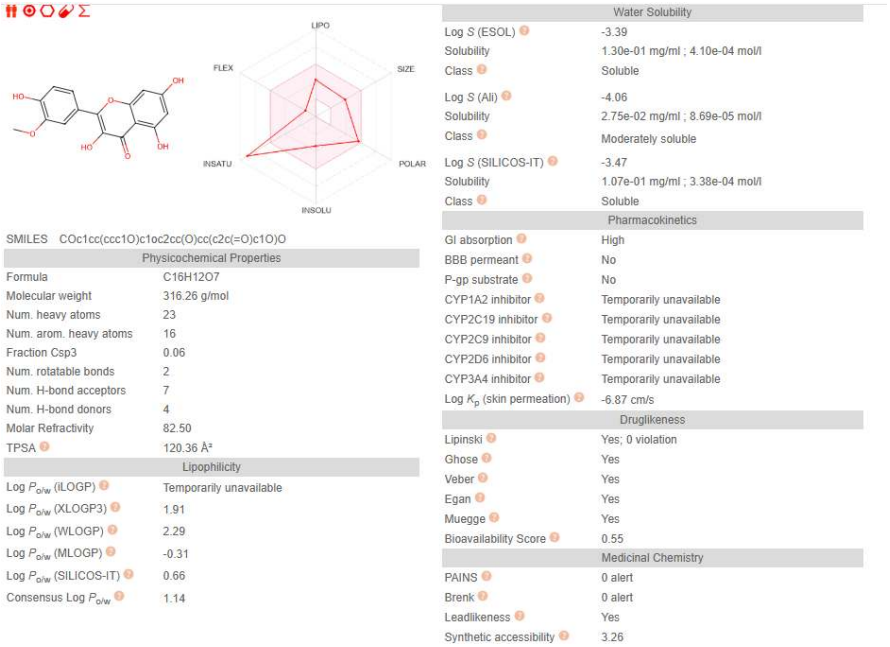


Figure 4.8: ***Consensus lipophilicity (Log Po/w) and aqueous solubility (Log S, ESOL model) for the three flavonols.***



SMILES COc1cc(ccc1O)c1oc2cc(O)cc(c2c(=O)c1O)O

Physicochemical Properties	
Formula	C16H12O7
Molecular weight	316.26 g/mol
Num. heavy atoms	23
Num. arom. heavy atoms	16
Fraction Csp3	0.06
Num. rotatable bonds	2
Num. H-bond acceptors	7
Num. H-bond donors	4
Molar Refractivity	82.50
TPSA	120.36 Å²
Lipophilicity	
Log $P_{o/w}$ (iLOGP)	Temporarily unavailable
Log $P_{o/w}$ (XLOGP3)	1.91
Log $P_{o/w}$ (WLOGP)	2.29
Log $P_{o/w}$ (MLOGP)	-0.31
Log $P_{o/w}$ (SILICOS-IT)	0.66
Consensus Log $P_{o/w}$	1.14

Water Solubility	
Log *S* (ESOL)	-3.39
Solubility	1.30e-01 mg/ml ; 4.10e-04 mol/l
Class	Soluble
Log *S* (Ali)	-4.06
Solubility	2.75e-02 mg/ml ; 8.69e-05 mol/l
Class	Moderately soluble
Log *S* (SILICOS-IT)	-3.47
Solubility	1.07e-01 mg/ml ; 3.38e-04 mol/l
Class	Soluble
Pharmacokinetics	
GI absorption	High
BBB permeant	No
P-gp substrate	No
CYP1A2 inhibitor	Temporarily unavailable
CYP2C19 inhibitor	Temporarily unavailable
CYP2C9 inhibitor	Temporarily unavailable
CYP2D6 inhibitor	Temporarily unavailable
CYP3A4 inhibitor	Temporarily unavailable
Log K_p (skin permeation)	-6.87 cm/s
Druglikeness	
Lipinski	Yes; 0 violation
Ghose	Yes
Veber	Yes
Egan	Yes
Muegge	Yes
Bioavailability Score	0.55
Medicinal Chemistry	
PAINS	0 alert
Brenk	0 alert
Leadlikeness	Yes
Synthetic accessibility	3.26

Figure 4.9: ***SwissADME physicochemical, pharmacokinetic, and drug-likeness output panel for isorhamnetin.***

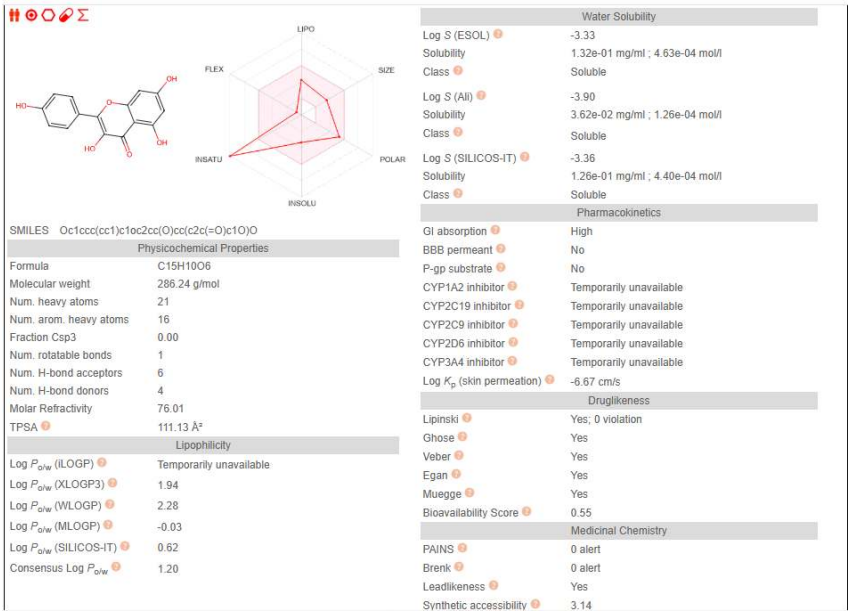


Figure 4.10: ***SwissADME physicochemical, pharmacokinetic, and drug-likeness output panel for kaempferol.***

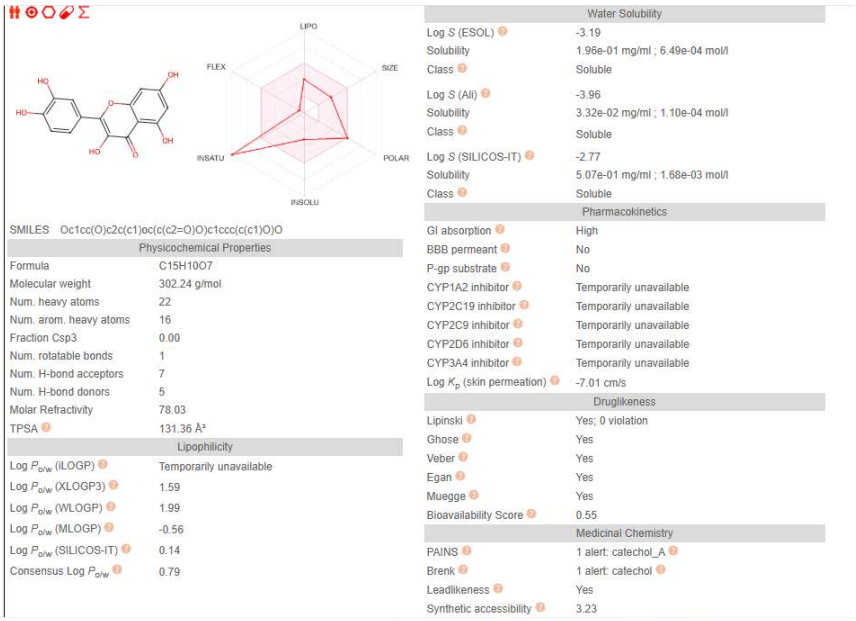


Figure 4.11: ***SwissADME physicochemical, pharmacokinetic, and drug-likeness output panel for quercetin, showing the catechol_A PAINS alert unique to this compound.***

4.7 Predicted ADMET Behaviour

All three compounds were predicted to have high gastrointestinal absorption, to be non-permeant across the blood-brain barrier, and to be non-substrates of P-glycoprotein, a combination that suggests reasonable oral exposure with limited central nervous system penetration — a broadly favourable profile if the intended pharmacological effect is peripheral (for example, renal or hepatic) rather than central. Skin permeation coefficients (Log Kp) were closely clustered in the range of -6.67 to -7.01 cm/s, indicating low intrinsic dermal permeability for all three, consistent with their polyphenolic, multiply hydroxylated structures. Cytochrome P450 isoform interaction predictions were reported as "temporarily unavailable" within the SwissADME output retrieved for this study and are therefore not presented as SwissADME-derived findings in this manuscript; the corresponding cytochrome P450 metabolism liability reported in Section 4.8 instead

derives from the independent ProTox-3.0 platform, whose CYP predictions are summarised below.

4.8 Toxicological Profile

The three flavonols diverged most sharply, of any dataset examined in this study, in their ProTox-3.0-predicted acute oral toxicity classification. Quercetin was assigned to toxicity Class 3 with a predicted LD50 of 159 mg/kg, whereas both kaempferol (LD50 3919 mg/kg) and isorhamnetin (LD50 5000 mg/kg) were assigned to the substantially safer Class 5. Under the Globally Harmonised System-style scale used by ProTox, Class 3 corresponds to compounds considered toxic if swallowed, while Class 5 corresponds to compounds that may be harmful if swallowed, a materially different risk categorisation for what are, structurally, two very closely related molecules differing by a single hydroxyl group. It is important to note that ProTox itself reported markedly different confidence metrics for the three predictions: the quercetin prediction was returned with 100% average similarity to its nearest training-set analogue and 100% prediction accuracy (indicating that quercetin itself, or a near-identical structure, is directly represented in the ProTox training data), whereas the kaempferol and isorhamnetin predictions were returned with lower average similarity (82.46% and 87.48% respectively) and a correspondingly lower general model accuracy (70.97%), meaning the quercetin toxicity classification should be regarded as the most directly evidenced of the three, while the kaempferol and isorhamnetin classifications rely more heavily on extrapolation from structurally related but non-identical training compounds.

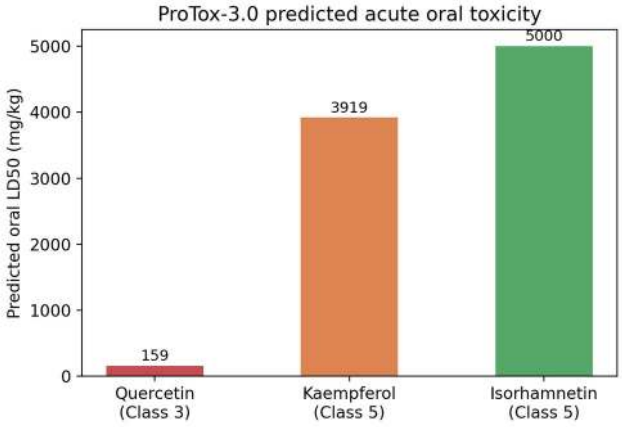


Figure 4.12: ***ProTox-3.0-predicted oral LD50 and toxicity class for quercetin, kaempferol, and isorhamnetin.***

Beyond acute lethality, all three compounds shared an almost identical qualitative organ- and pathway-toxicity fingerprint: all were predicted active (i.e., a positive toxicological signal) for nephrotoxicity, respiratory toxicity, blood-brain barrier disruption, and nutritional toxicity, and all were predicted active across the Aryl hydrocarbon Receptor (AhR), Estrogen Receptor Alpha (ERα), ER-LBD, and Mitochondrial Membrane Potential (MMP) endpoints, indicating a shared pattern of moderate general reactivity typical of polyphenolic natural products. Quercetin alone was flagged active for carcinogenicity and mutagenicity, both endpoints for which it returned the very high (100%/100%) similarity and accuracy metrics noted above, giving these two particular quercetin flags relatively strong evidentiary weight within the present dataset. Isorhamnetin alone was flagged active for cardiotoxicity, immunotoxicity, and ATAD5 (a DNA-damage-response stress marker), while kaempferol's overall active-endpoint count was the lowest of the three compounds. Cytochrome P450 interaction predictions from Pro Tox showed quercetin as the most broadly cytochrome-interacting of the three (active for CYP1A2, CYP2C19, and CYP2C9), kaempferol active for four of six isoforms tested (CYP1A2, CYP2C19, CYP2C9, CYP2D6), and isorhamnetin active for three (CYP1A2, CYP2C19, CYP2C9), with none of the three predicted to interact with CYP3A4 or CYP2E1.

Table 4.4 (below) summarises the full ProTox-3.0 model report for all thirty-one shared classification

endpoints across the three compounds; probabilities are given in parentheses, and "Active" denotes a positive toxicological signal for that endpoint.

Table 4.4: ***Comparative ProTox-3.0 toxicity model report for kaempferol, isorhamnetin, and quercetin. Full single-compound reports, including toxicity-target pharmacophore data, are reproduced in Appendix B.***

Endpoint	Kaempferol	Isorhamnetin	Quercetin
Hepatotoxicity	Inactive (0.68)	Inactive (0.72)	Inactive (0.69)
Neurotoxicity	Inactive (0.89)	Inactive (0.88)	Inactive (0.89)
Nephrotoxicity	Active (0.62)	Active (0.64)	Active (0.62)
Respiratory toxicity	Active (0.83)	Active (0.85)	Active (0.83)
Cardiotoxicity	Inactive (0.91)	Active (0.82)	Inactive (0.99)
Carcinogenicity	Inactive (0.72)	Inactive (0.68)	Active (0.68)
Immunotoxicity	Inactive (0.96)	Active (0.58)	Inactive (0.87)
Mutagenicity	Inactive (0.52)	Inactive (0.94)	Active (0.51)
Cytotoxicity	Inactive (0.98)	Inactive (0.95)	Inactive (0.99)
BBB-barrier	Active (0.57)	Active (0.52)	Active (0.53)
Ecotoxicity	Inactive (0.51)	Inactive (0.55)	Inactive (0.53)
Clinical toxicity	Inactive (0.54)	Inactive (0.50)	Inactive (0.53)
Nutritional toxicity	Active (0.66)	Active (0.57)	Active (0.63)
AhR (nr_ahr)	Active (1.00)	Active (0.97)	Active (0.91)
Androgen receptor (AR)	Inactive (0.99)	Inactive (1.00)	Inactive (0.99)
AR-LBD	Inactive (0.99)	Inactive (0.99)	Inactive (0.97)
Aromatase	Active (0.96)	Active (0.58)	Inactive (0.91)
Estrogen receptor α (ER)	Active (1.00)	Active (0.88)	Active (0.87)
ER-LBD	Active (0.95)	Active (0.89)	Active (0.95)
PPAR-γ	Inactive (0.95)	Inactive (0.95)	Inactive (0.98)
NRF2/ARE (sr_are)	Inactive (0.99)	Inactive (0.96)	Inactive (0.99)
HSE (sr_hse)	Inactive (0.99)	Inactive (0.96)	Inactive (0.99)
Mitochondrial membrane potential	Active (1.00)	Active (0.92)	Active (1.00)
p53	Inactive (0.92)	Inactive (0.86)	Inactive (0.97)
ATAD5	Inactive (0.92)	Active (0.65)	Inactive (0.99)
CYP1A2	Active (0.93)	Active (0.91)	Active (1.00)
CYP2C19	Active (0.65)	Active (0.83)	Active (0.77)
CYP2C9	Active (0.66)	Active (0.63)	Active (0.99)
CYP2D6	Active (0.62)	Inactive (0.56)	Inactive (0.85)
CYP3A4	Inactive (0.65)	Inactive (0.94)	Inactive (0.79)
CYP2E1	Inactive (1.00)	Inactive (0.99)	Inactive (1.00)

4.9 Predicted Protein Targets and Polypharmacology Landscape

SwissTargetPrediction did not, for any of the three compounds, return GAL3ST1 (or any other sulfotransferase family member) among its predicted targets. As explained in Section 2.5 and returned to at length in Section 5.5, this is an expected and mechanistically uninformative null result rather than evidence against GAL3ST1 engagement, because SwissTargetPrediction's algorithm can only nominate targets that are already represented, with an annotated small-molecule ligand, in its curated reference library, and GAL3ST1 currently has no such annotated ligand in the databases underlying that library. The tool's output is therefore reported and interpreted here purely as an off-target and polypharmacology risk profile, run in parallel with, and entirely independent of, the

GAL3ST1-directed docking analysis in Sections 4.2 to 4.5.

With that caveat established, the three flavonols showed substantial convergence in their top-ranked predicted targets. Carbonic anhydrase isoforms (particularly CA2, CA4, CA7, CA9, CA12, and CA14), cytochrome P450 1B1, the ATP-binding-cassette efflux transporter ABCG2, and the phosphoinositide kinases IP6K2 and IPMK appeared among the highest-probability (typically probability 1.00) predicted targets for all three compounds, indicating a shared, chemotype-driven affinity for these protein families that is a well-documented general feature of polyphenolic flavonoids rather than a property specific to any one of the three molecules studied here. Kaempferol and quercetin in fact shared five of their top-ten predicted targets outright (ABCG2, TRPC5, IP6K2, HDAC6, and NOX4), while isorhamnetin's top-ten list was comparatively enriched for carbonic anhydrase isoforms (CA2, CA4, CA7, CA12) and the nuclear receptor PPAR-gamma, sharing only IP6K2, CYP1B1, and XDH with kaempferol and essentially no top-ten overlap with quercetin. This pattern — the two unmethylated, more polar flavonols (kaempferol and quercetin) clustering together, with the O-methylated isorhamnetin showing a somewhat distinct off-target signature — mirrors, at the level of ligand-similarity-based target prediction, the same single-methylation structural distinction that was associated with isorhamnetin's improved Cavity 1 docking score in Section 4.4.

Table 4.5: Top Ten SwissTargetPrediction Hits for Kaempferol.

Predicted target	Target class	Probability
ABCG2	Primary active transporter	1.00
TRPC5	Voltage-gated ion channel	1.00
IP6K2	Transferase	1.00
HDAC6	Eraser	1.00
CISD1	Unclassified protein	1.00
NOX4	Oxidoreductase	1.00
IPMK	Transferase	1.00
CYP1B1	Cytochrome P450	1.00
RPS6KA3	Kinase	1.00
XDH	Oxidoreductase	1.00

Table 4.6: Top Ten Swiss TargetPrediction Hits for Isorhamnetin.

Predicted target	Target class	Probability
IP6K2	Transferase	1.00
IPMK	Transferase	1.00
CYP1B1	Cytochrome P450	1.00
XDH	Oxidoreductase	1.00
CA7	Lyase	1.00
PPARG	Nuclear receptor	1.00
CA4	Lyase	1.00
CA2	Lyase	1.00
CA12	Lyase	1.00
ELAVL3	Unclassified protein	0.988

Table 4.7: Top Ten Swiss Target Prediction Hits for Quercetin.

Predicted target	Target class	Probability
ABCG2	Primary active transporter	1.00
ALK	Kinase	1.00

Predicted target	Target class	Probability
CA14	Lyase	1.00
TRPC5	Voltage-gated ion channel	1.00
IP6K2	Transferase	1.00
HDAC6	Eraser	1.00
SPHK1	Enzyme	1.00
NOX4	Oxidoreductase	1.00
NEK6	Kinase	1.00
AURKB	Kinase	1.00

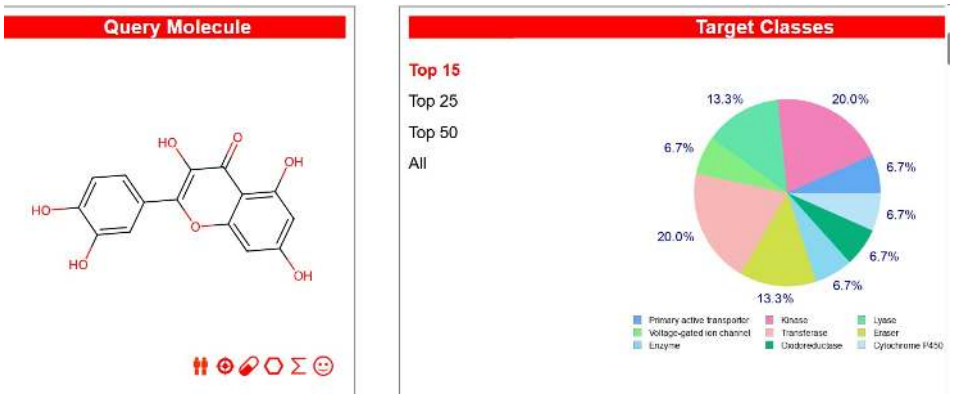


Figure 4.13: ***SwissTargetPrediction target-class distribution (top 15 predicted targets) for quercetin.***



Figure 4.14: ***SwissTargetPrediction target-class distribution (top 15 predicted targets) for kaempferol.***

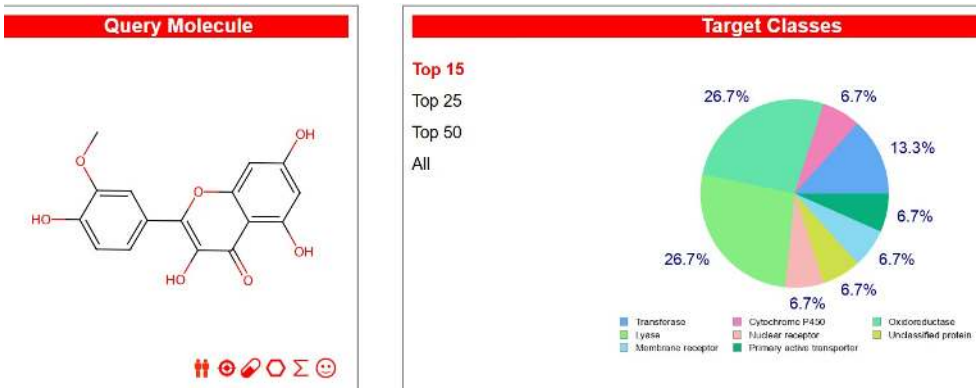


Figure 4.15: ***SwissTargetPrediction target-class distribution (top 15 predicted targets) for isorhamnetin.***

4.10 The Boltz Affinity Dataset

Compound_1 returned a Boltz predicted binding probability of 0.942, a predicted IC50 of 19.1 nanomolar, and a "strong binder" classification, while Compound_2 returned a predicted binding probability of 0.975, a predicted IC50 of 4.1 nanomolar, and likewise a "strong binder" classification, both against whatever receptor target that particular Boltz run was configured for.

DISCUSSION

5.1 Structural Rationale for a Shared Binding Cavity

The single most consistent finding of this study is that three independently docked, structurally related ligands converged, without any user-imposed constraint, on the same cavity of the GAL3ST1 model. Blind docking makes no a priori commitment to any particular site, so this convergence is not a foregone methodological conclusion but an emergent result: had Cavity 1 been geometrically or chemically unsuited to accommodate a 3-hydroxyflavone scaffold, at least one of the three ligands would have been expected to favour a different cavity, particularly given that isorhamnetin's additional methyl group and altered polarity, relative to quercetin, could plausibly have shifted its preferred binding mode. That all three ligands instead ranked Cavity 1 first, and did so with docking scores clustered within a narrow 0.3 kcal/mol range, is best explained by Cavity 1 possessing both sufficient volume (roughly thirteen-fold larger than any other detected cavity) and a sufficiently varied chemical surface (a mixture of basic, aromatic, and polar residues) to comfortably and near-interchangeably accommodate all three closely related flavonol chemotypes.

5.2 Structure-Activity Relationships Among the Three Flavonols

Read together, the docking, physicochemical, and toxicological results outline a coherent, if modest, structure-activity narrative across the quercetin-kaempferol-isorhamnetin series. Removing quercetin's second B-ring hydroxyl to give kaempferol left the Cavity 1 docking score essentially unchanged (-9.6 kcal/mol for both), implying that this particular hydroxyl does not make an indispensable, score-determining contact within Cavity 1, even though it clearly does contribute to quercetin's distinguishing catechol-related structural alerts and possibly to its higher acute toxicity classification (Section 5.4). Conversely, methylating one of quercetin's two catechol hydroxyls to give isorhamnetin was associated with the single best Cavity 1 score of the three compounds (-9.9 kcal/mol) and with an additional peripheral contact surface (Val34-Leu37, Tyr204-Asp225) not engaged by either quercetin or kaempferol, suggesting that the added methyl group and the resulting shift in overall molecular shape allow isorhamnetin to reach slightly deeper into, or further around, the Cavity 1 pocket than its unmethylated relatives. Taken together, these observations suggest that Cavity 1 tolerates B-ring hydroxylation state reasonably well but rewards the added steric bulk and altered polarity of O-methylation with a marginal binding-score improvement, a pattern of practical relevance should this series be advanced toward the synthesis of close structural analogues.

5.3 Druggability and Medicinal Chemistry Considerations

From a strictly medicinal-chemistry standpoint, all three flavonols are attractive as chemical starting points precisely because they combine reasonable predicted target engagement (a consensus, favourable-scoring cavity) with essentially ideal drug-likeness metrics: low molecular weight, low-to-moderate lipophilicity, few rotatable bonds, and full compliance with every rule-based filter applied. Quercetin's catechol-related PAINS and Brenk alerts are the one clear liability distinguishing it from its two analogues, and these alerts specifically flag a chemical mechanism — auto-oxidation to a reactive ortho-quinone and non-specific metal chelation — that is well documented in the broader natural-product literature as a source of assay artefacts and non-specific reactivity for catechol-bearing polyphenols including quercetin itself. This does not mean quercetin's docking result is invalid, but it does mean that any future biochemical or cell-based assay used to validate the GAL3ST1 docking prediction for quercetin specifically should be designed with appropriate counter-screens (for example, redox-cycling and thiol-reactivity controls) to distinguish genuine,

specific enzyme engagement from these well-characterised catechol-driven artefact mechanisms; kaempferol and isorhamnetin, lacking a free catechol, would not require the same degree of counter-screening.

5.4 Safety Margins: Interpreting the Divergent ProTox Classifications

The nearly thirty-fold difference in predicted LD50 between quercetin (159 mg/kg) and isorhamnetin (5000 mg/kg) is the most striking single number generated by this study, and it warrants careful, non-alarmist interpretation. As noted in Section 4.8, the quercetin prediction carries a 100% training-set similarity and 100% reported accuracy, meaning ProTox is very likely drawing on data for quercetin itself (or a near-identical reference compound) rather than extrapolating across chemical space; this lends the quercetin classification real evidentiary weight rather than dismissing it as a modelling artefact. At the same time, quercetin has a long history of dietary consumption at gram-per-day intakes in some supplement contexts without the kind of overt acute toxicity that a Class 3 (toxic if swallowed) classification might, out of context, be taken to imply; the apparent tension is most plausibly resolved by recognising that ProTox's LD50 model is typically trained predominantly on parenteral or high-dose rodent acute-toxicity study data, which does not necessarily map in a simple, linear way onto chronic, low-dose human dietary exposure, and that a computationally predicted acute LD50 is, in any case, a hazard signal to be followed up experimentally rather than a definitive statement about real-world human safety at typical exposure levels. The comparatively benign Class 5 classifications for kaempferol and isorhamnetin, while resting on lower training-set similarity and therefore somewhat greater extrapolation uncertainty, are nonetheless directionally consistent with the generally favourable safety record reported for these two compounds in the wider preclinical literature (Section 2.3), and on balance this study's toxicity data favour prioritising kaempferol and isorhamnetin over quercetin for any downstream experimental work specifically targeting GAL3ST1.

5.5 Off-Target Signals and What Swiss Target Prediction Does and Does Not Tell Us

It is essential to state plainly that the complete absence of GAL3ST1 from every Swiss Target Prediction output in this study is not evidence against the docking findings reported in Sections 4.2 through 4.5; it is a direct and fully expected consequence of the algorithm's reliance on chemical similarity to a curated library of molecules with experimentally annotated targets, a library that, at present, contains no small molecule annotated as a GAL3ST1 ligand for any of the three flavonols (or any other compound) to be judged similar to. In other words, Swiss Target Prediction and the CB-Dock2 blind-docking analysis are answering two entirely different, complementary questions: the former asks "which already-characterised targets does this molecule chemically resemble known ligands for," while the latter asks "does this specific molecule fit geometrically and chemically into a cavity on this specific, admittedly uncharacterised, receptor." A structure-based approach such as blind docking is, by design, the only one of the two methods capable of generating any hypothesis at all about an emerging, ligand-poor target such as GAL3ST1, and this asymmetry is itself a useful, generalisable methodological lesson for other researchers approaching similarly under-annotated enzymes.

That said, the Swiss Target Prediction output remains genuinely informative as an independent off-target and poly pharmacology risk assessment, run in parallel with, rather than in place of, the GAL3ST1-specific docking work. The recurring appearance of carbonic anhydrase isoforms, cytochrome P450 1B1, and the ABCG2 efflux transporter across all three compounds' target-prediction lists is consistent with a large existing pharmacological literature documenting flavonoid interactions with each of these protein families, and it flags carbonic anhydrase inhibition and ABCG2-mediated efflux interactions specifically as the polypharmacology effects most likely to be observed, and that should therefore be explicitly controlled for, in any future cell-based assay designed to attribute a phenotypic effect specifically to GAL3ST1 inhibition rather than to one of these better-characterised off-target activities.

5.6 Extended Consideration of the Ovarian Cancer and Metabolic Disease Contexts

While the renal cell carcinoma literature offers the most mechanistically developed rationale for GAL3ST1-directed intervention (Sections 1.3, 2.2, and 2.5), the reported elevation of GAL3ST1 transcript levels in ovarian epithelial carcinoma deserves separate comment, since ovarian cancer's persistently poor five-year survival relative to many other solid tumours, driven substantially by late-stage diagnosis and the frequent emergence of platinum resistance, means that any additional, mechanistically distinct vulnerability specific to this tumour type would be of considerable clinical interest even at a very early, exploratory stage of characterisation. No study to date has, to the authors' knowledge, mechanistically dissected why GAL3ST1 expression rises in ovarian epithelial carcinoma or whether the sulfatide-platelet immune-evasion mechanism described for renal cell carcinoma also operates in the ovarian context; this represents a specific, well-defined open question that the computational findings of the present study cannot address but that would meaningfully strengthen the translational case for GAL3ST1-directed flavonol candidates such as kaempferol and isorhamnetin if resolved in the affirmative by future mechanistic work.

The metabolic-disease associations reviewed in Sections 1.3 and 2.2 — the linkage between GAL3ST1 genetic variation and insulin resistance, and the dynamic regulation of hepatic Gal3st1 expression during acute kidney injury — point toward a second, quite distinct potential application distinct from oncology, namely modulation of systemic sulfatide levels in metabolic disease. Because quercetin, kaempferol, and isorhamnetin are all dietary constituents already consumed as part of habitual fruit and vegetable intake, any effect of chronic dietary flavonol consumption on circulating sulfatide levels or insulin sensitivity via GAL3ST1 modulation would be, in principle, amenable to retrospective analysis in existing dietary-intake epidemiological cohorts, offering a comparatively low-cost avenue for early hypothesis testing that does not require de novo drug development infrastructure.

5.7 Placing the Findings in the Context of GAL3ST1 Disease Biology

If the Cavity 1 binding predicted here for all three flavonols were to be confirmed biochemically as genuine enzyme inhibition, the disease contexts reviewed in Section 1.3 suggest renal cell carcinoma as the most immediately compelling application, given the direct mechanistic link already established in the literature between GAL3ST1-driven sulfatide accumulation, tumour cell-platelet binding, and immune evasion in that specific tumour type; a small molecule capable of even partial GAL3ST1 inhibition could, in principle, be tested for its ability to blunt this axis. At the same time, the essential physiological roles of sulfatide in myelin structure and in spermatogenesis (Section 1.3) mean that any translational programme building on these findings would need, from an early stage, to characterise the degree of GAL3ST1 inhibition achievable at exposures well below any threshold expected to compromise baseline nervous-system or reproductive-tissue sulfatide levels — a therapeutic-window question that lies well beyond the scope of the present purely computational study but that should be flagged explicitly for any research group choosing to build upon it.

5.8 Limitations

The principal limitations of this study, several of which were flagged prospectively in Section 3.9, are summarised here for completeness:

- The GAL3ST1 receptor model has no experimental structural validation and no independently confirmed active-site annotation; all conclusions about the identity and druggability of Cavity 1 are therefore model-dependent hypotheses rather than experimentally established facts.
- A single physics-based docking scoring function (AutoDock Vina, as implemented within CB-Dock2) was used throughout; no consensus docking, molecular dynamics refinement, or free-energy perturbation calculation was performed to corroborate the reported binding scores.
- The identity of the third, unlabelled CurPockets data file as corresponding specifically to isorhamnetin (rather than to some other compound) was inferred by the process of elimination against the

two explicitly labelled quercetin and kaempferol files, and, while internally consistent with the isorhamnetin-specific ProTox and SwissADME datasets supplied alongside it, this attribution could not be independently confirmed from the file's own metadata.

- SwissADME, ProTox-3.0, and SwissTargetPrediction are ligand-based, statistical, and similarity-driven prediction platforms; none of them incorporates any information about GAL3ST1 specifically, and their outputs describe general pharmacokinetic, toxicological, and off-target properties of the three compounds rather than any GAL3ST1-specific behaviour.
- The ProTox-3.0 acute-toxicity predictions for kaempferol and isorhamnetin were returned with lower training-set similarity than the quercetin prediction, and should accordingly be treated with somewhat greater caution pending experimental confirmation.
- No in vitro or in vivo experimental data of any kind were generated or reviewed for this manuscript; every finding reported here is a computational prediction awaiting biochemical validation.

5.9 Comparison with Other Computational Studies of Flavonoid-Enzyme Interactions

The docking scores obtained here for quercetin, kaempferol, and isorhamnetin against the GAL3ST1 Cavity 1 site (-9.6 to -9.9 kcal/mol) fall within the range commonly reported in the wider computational literature for flavonoid docking against a variety of established enzyme targets, including kinases, carbonic anhydrases, and viral proteases, where flavonol scaffolds frequently return AutoDock Vina scores in the -7 to -10 kcal/mol band. This places the present GAL3ST1 results well within a plausible, literature-consistent range for genuine, if moderate, target engagement, rather than in the much more negative range (below approximately -11 to -12 kcal/mol) that would raise concern about scoring-function artefacts or unrealistically deep burial of the ligand. At the same time, because no other published study has, to the authors' knowledge, docked any small molecule against GAL3ST1 specifically, no direct external benchmark exists against which to calibrate whether -9.6 to -9.9 kcal/mol against this particular receptor model corresponds to a biologically meaningful or merely a computationally plausible interaction; this is precisely the gap that the experimental validation work recommended in Section 7 is intended to close.

5.10 Broader Implications for Natural-Product-Based Discovery Against Understudied Targets

Beyond its specific findings for GAL3ST1, this study illustrates a broader, generalisable workflow for approaching the substantial fraction of the human proteome that, like GAL3ST1, is implicated in disease by genetic, transcriptomic, or correlative evidence but has never been the subject of a dedicated small-molecule discovery campaign and lacks any experimentally resolved structure or annotated ligand-binding site. For such targets, the combination of (i) a computationally predicted receptor model, (ii) cavity-detection-guided blind docking against a small, structurally related, well-characterised ligand series, and (iii) parallel, independent ADMET, toxicity, and off-target profiling of that same ligand series offers a tractable, resource-light route to a first, testable structure-activity hypothesis, without requiring the receptor-specific experimental investment (crystallography, mutagenesis, or biochemical assay development) that would otherwise need to precede any computational work at all. The explicit, repeated acknowledgement throughout this manuscript that SwissTargetPrediction cannot, by construction, nominate a ligand-poor target such as GAL3ST1 (Section 5.5) is offered as a specific methodological lesson of potentially wider relevance to other research groups working on comparably understudied enzymes drawn from genome-wide association, transcriptomic, or hypoxia-response literature of the kind reviewed in Section 2.2.

5.11 Integrated Prioritisation Scorecard

Table 5.1 consolidates the principal quantitative findings from Sections 4.2 through 4.9 into a single side-by-side scorecard, intended to make the overall prioritisation argument developed across this Discussion immediately auditable against the underlying numbers rather than requiring the reader to cross-reference several earlier tables.

Table 5.1: ***Integrated scorecard consolidating the docking, physicochemical, and toxicological findings of Sections 4.2-4.9 used to support the prioritisation argument developed in Sections 5.2-5.4.***

Metric	Quercetin	Kaempferol	Isorhamnetin
Cavity 1 Vina score (kcal/mol)	-9.6	-9.6	-9.9
Molecular weight (g/mol)	302.24	286.24	316.26
Consensus Log Po/w	0.79	1.20	1.14
TPSA (Å²)	131.36	111.13	120.36
Lipinski / Veber / Egan / Muegge violations	0 / 0 / 0 / 0	0 / 0 / 0 / 0	0 / 0 / 0 / 0
PAINS alerts	1 (catechol_A)	0	0
Brenk alerts	1 (catechol)	0	0
Predicted oral LD50 (mg/kg)	159	3919	5000
ProTox toxicity class	Class 3	Class 5	Class 5
ProTox training-set similarity	100%	82.46%	87.48%
GI absorption	High	High	High
BBB permeant	No	No	No
Overall prioritisation (this study)	Third	First (tied)	First (tied)

CONCLUSION

This study applied an unbiased, cavity-detection-guided blind docking strategy, necessitated by the complete absence of experimental structural or active-site data for cerebroside sulfotransferase (GAL3ST1), to evaluate three closely related dietary flavonols — quercetin, kaempferol, and isorhamnetin — as candidate small-molecule ligands of this emerging oncology and metabolic-disease target. All three compounds independently converged on the same large, chemically mixed cavity of the GAL3ST1 model, achieving docking scores between -9.6 and -9.9 kcal/mol and sharing a consensus contact-residue fingerprint that included multiple basic, aromatic, and polar residues consistent, at least in principle, with the anion-binding character expected of a sulfotransferase catalytic groove. All three compounds additionally satisfied every standard rule-based drug-likeness filter and were predicted to have high oral absorption with low central nervous system exposure. They diverged most sharply in predicted acute toxicity, with quercetin classified as substantially more hazardous (Class 3, LD50 159 mg/kg) than either kaempferol or isorhamnetin (Class 5, LD50 3919 and 5000 mg/kg respectively), a difference reinforced by quercetin's unique catechol-related PAINS and Brenk structural alerts. On the balance of the computational evidence assembled here, kaempferol and isorhamnetin emerge as the more favourable candidates of the three for downstream experimental prioritisation, while quercetin's docking behaviour, though equally favourable in binding-score terms, is best interpreted alongside appropriate counter-screens for catechol-driven, target-independent reactivity. More broadly, this work illustrates how a combination of blind docking and standard in silico ADMET, toxicity, and target-prediction tools can generate a coherent, testable structure-activity and safety hypothesis for a target as experimentally under-characterised as GAL3ST1, and it is offered as a starting point for, rather than a substitute for, the biochemical validation that any such computational hypothesis ultimately requires.

7. Future Directions

Building directly on the findings and limitations documented above, the following lines of work are recommended as immediate next steps:

- Experimental validation of GAL3ST1 inhibition by quercetin, kaempferol, and isorhamnetin using a recombinant or cell-based sulfotransferase activity assay, ideally including a catechol-reactivity counter-screen specific to the quercetin arm of the series.
- Molecular dynamics simulation of the top-ranked Cavity 1 complexes to assess pose stability over time and to estimate binding free energies using

MM-GBSA/MM-PBSA methods, addressing the single-scoring-function limitation noted in Section 5.7.

- Generation, and independent structural validation where feasible (for example by comparison against any future experimentally solved GAL3ST1 structure or high-confidence deep-learning model with per-residue confidence scores), of an improved receptor model to test the robustness of the Cavity 1 assignment reported here.
- A confirmatory, correctly targeted Boltz (or equivalent deep-learning) affinity prediction run specifically on quercetin, kaempferol, and isorhamnetin against the GAL3ST1 model, replacing the chemically unrelated dataset noted in Sections 3.8 and 4.10.
- Design and synthesis of a small, focused series of isorhamnetin-like O-methylated flavonol analogues to test whether the peripheral Cavity 1 subpocket identified in Section 4.4 can be further exploited to improve binding score and selectivity.
- Cell-based phenotypic assays in renal cell carcinoma models, incorporating the carbonic anhydrase and ABCG2 off-target controls indicated by the SwissTargetPrediction analysis (Section 5.5), to test whether any observed phenotypic effect is attributable specifically to GAL3ST1 modulation.
- Direct CST/GAL3ST1 enzyme-inhibition assays (recombinant or membrane-preparation format) for the selected Clitoria ternatea flavonol constituents, together with cell-based sulfatide-estimation studies (for example by LC-MS/MS or immunostaining in an oligodendrocyte-lineage or patient-derived fibroblast model) to determine whether predicted binding translates into any measurable change in sulfatide level — the minimum experimental step required before these computationally prioritised compounds could be described as candidate substrate-reduction agents.
- Dedicated pharmacokinetic and blood-brain-barrier permeability evaluation (in vitro PAMPA-BBB or MDCK-MDR1 assay, followed by in vivo brain-to-plasma ratio studies) specific to the MLD application, since standard SwissADME BBB flags reported in Section 4.7 are general-purpose predictions and are not sufficient, on their own, to support any CNS-directed therapeutic claim.

7.1 Future Formulation Strategy for CNS-Directed Delivery

Even if downstream assays were to support a genuine CST-inhibitory effect for one or more of the selected Clitoria ternatea flavonol constituents, their plain, unformulated forms would remain poorly suited to a central-nervous-system indication such as MLD, owing to low intrinsic aqueous solubility, extensive first-pass glucuronidation/sulfation, and predicted low blood-brain-barrier permeability (Section 4.7). Any realistic translational pathway would therefore need to run in parallel with a dedicated formulation-development programme, and the following directions are proposed purely as forward-looking, hypothesis-level strategy rather than as work undertaken in the present study:

- Nanoparticulate carriers (PLGA or lipid-polymer hybrid nanoparticles, solid-lipid nanoparticles, or surface-functionalised liposomes) to improve aqueous dispersibility, protect the flavonol core from rapid first-pass metabolism, and provide a platform for eventual ligand-mediated (for example transferrin- or glucose-receptor-targeted) brain delivery.
- Intranasal (nose-to-brain) delivery systems, including mucoadhesive nanoemulsions or in-situ gelling formulations, as a non-invasive route capable of partially bypassing the blood-brain barrier via the olfactory and trigeminal pathways, a strategy already explored for other lysosomal and neurodegenerative indications.
- Cyclodextrin inclusion complexation or amorphous solid-dispersion approaches as lower-complexity, comparatively low-cost solubility-enhancement strategies, consistent with the cost-reduction rationale underpinning the choice of a dietary flavonol source over a novel synthetic chemical entity (Section 1.3.1).
- Standardised Clitoria ternatea extract or fraction-based formulations (rather than single-isolate compound products) as a pragmatic, lower-cost

intermediate strategy, together with formal extract standardisation (HPLC/HPTLC fingerprinting against the flavonol constituents modelled here) to ensure batch-to-batch consistency before any pharmacokinetic or efficacy testing.

All of the above remain speculative, formulation-stage proposals contingent on positive outcomes at every preceding experimental step (Sections 3.10 and 7); none should be read as evidence that a deliverable CNS formulation currently exists or that the flavonol constituents examined here have been shown to reach the brain in an active form.

INTRODUCTION

1.1 Glycosphingolipids and Sulfatide Biology

Glycosphingolipids are amphipathic membrane lipids built from a ceramide backbone linked to one or more sugar residues, and they are indispensable structural and signalling components of the plasma membrane and of specialised organelles such as the Golgi apparatus and myelin sheath. Among the simplest of the glycosphingolipids is galactosylceramide (GalCer), formed by the transfer of a single galactose unit onto ceramide. Galactosylceramide is itself the immediate biosynthetic precursor of sulfatide (3-O-sulfogalactosylceramide), one of the most abundant acidic glycosphingolipids of the vertebrate nervous system, where it constitutes a substantial fraction of the myelin sheath's outer leaflet and contributes to axon-glia adhesion, ion channel clustering at the paranode, and the general biophysical stability of compact myelin.

Beyond the nervous system, sulfatides are expressed at lower but functionally important levels in epithelial tissues of the kidney, gastrointestinal tract, and reproductive organs, and they participate in processes as diverse as platelet activation, complement regulation, insulin secretion, and the adhesion of circulating tumour cells to the vascular endothelium. Because sulfatide levels are governed almost entirely by the balance between a single sulfation enzyme and a single desulfating enzyme, disturbances at either step have outsized physiological consequences. Loss-of-function mutations in arylsulfatase A, the lysosomal enzyme that removes the sulfate group from sulfatide, cause the fatal, sulfatide-storage disease metachromatic leukodystrophy, illustrating how dependent normal cellular physiology is on tightly regulated sulfatide turnover. The enzyme that performs the opposite, sulfation-adding reaction is the central focus of the present study.

1.2 Cerebroside Sulfotransferase (GAL3ST1): Gene, Protein, and Catalytic Function

Cerebroside sulfotransferase, more formally galactosylceramide sulfotransferase and officially designated by the gene symbol GAL3ST1 (galactose-3-O-sulfotransferase 1; EC 2.8.2.11; UniProt accession Q99999; NCBI Gene ID 9514; historically abbreviated CST), catalyses the transfer of a sulfonate group from the universal sulfate donor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) onto the 3-hydroxyl of the galactose ring of galactosylceramide, yielding sulfatide and the spent co-substrate adenosine 3',5'-bisphosphate. This single reaction represents the final, committed step of sulfatide biosynthesis, making GAL3ST1 the rate-limiting and pharmacologically most tractable node for pharmacological modulation of cellular sulfatide content.

The human GAL3ST1 gene resides at chromosome 22q12.2 and encodes a 423-amino-acid polypeptide with a predicted type II single-pass transmembrane topology: a short cytosolic N-terminal tail, a single transmembrane anchor, and a large luminal catalytic domain that resides within the trans-Golgi network, oriented to act upon glycolipid substrates presented on the luminal leaflet of the Golgi membrane. Sequence analyses place GAL3ST1 in a distinct sulfotransferase sub-family that is not homologous to either the classical cytosolic sulfotransferases (which sulfonate small-molecule xenobiotics, hormones, and neurotransmitters) or the majority of other Golgi-resident carbohydrate sulfotransferases, and, consistent with this, GAL3ST1 possesses two potential N-linked glycosylation sites and a catalytic fold that has not, to date, been solved by X-ray crystallography or cryo-electron microscopy. In the absence of an experimental structure, computational models (for example, template-free deep-learning structure predictions) are the only currently available route to a three-dimensional representation of the enzyme, and — critically for the methodology adopted in this study — none of the

available models carries a curated, literature-validated annotation of the PAPS-binding or acceptor-binding sub-sites, since these have never been mapped experimentally for this particular sulfotransferase. This absence of a validated domain-level active-site description is the direct reason a blind, rather than site-directed, docking strategy was required for the present work, and it is discussed at length in Section 3.

1.2 GAL3ST1 in Human Disease

Interest in GAL3ST1 as a pharmacological target stems from a growing body of correlative and mechanistic evidence linking sulfatide overproduction to malignant and metabolic phenotypes. In clear-cell renal cell carcinoma, hypoxia-inducible factor 1 has been shown to transcriptionally upregulate GAL3ST1, driving a hypoxia-inducible HIF1-GAL3ST1-sulfatide axis that increases tumour-cell surface sulfatide and, through enhanced tumour cell-platelet binding, promotes immune evasion; independent clinical cohorts have further reported that elevated GAL3ST1/Gal3ST1 expression carries prognostic significance in renal cell carcinoma. GAL3ST1 transcript levels are likewise elevated in ovarian epithelial carcinoma, and the encoded enzymatic activity is increased in renal cell carcinoma tissue relative to normal parenchyma, reinforcing sulfatide accumulation as a recurring feature of the malignant phenotype in at least two solid-tumour lineages.

Outside oncology, genetic variation at the GAL3ST1 locus has been associated with exercise-modified insulin resistance and with type 2 diabetes risk, and rodent studies have shown that hepatic Gal3st1 expression is acutely downregulated in protein-overload nephropathy-induced acute kidney injury, with a corresponding fall in hepatic and serum sulfatide, linking this single enzyme to whole-body lipid and metabolic homeostasis well beyond the nervous system. In parallel, genetic ablation of Gal3st1 in mice — which abolishes sulfatide synthesis entirely — produces defects in paranodal junction formation and in spermatogenesis, underscoring that sulfatide, and by extension the enzyme that makes it, performs genuinely essential physiological roles that any pharmacological strategy directed at GAL3ST1 would need to respect, particularly with regard to the nervous system and reproductive tissue. Taken together, these observations position GAL3ST1 simultaneously as a plausible small-molecule target for oncology and metabolic disease and as an enzyme whose inhibition would need to be carefully titrated to avoid disturbing baseline myelin and reproductive physiology — a tension that is returned to in the Discussion.

1.3.1 Relevance to Metachromatic Leukodystrophy: A Substrate-Reduction Rationale

Metachromatic leukodystrophy (MLD) is a rare, autosomal-recessive lysosomal storage disorder caused principally by deficient activity of arylsulfatase A (ARSA), the lysosomal enzyme responsible for the stepwise desulfation and catabolism of sulfatide (3-O-sulfogalactosylceramide). Loss of ARSA activity results in progressive lysosomal and extralysosomal accumulation of sulfatide within oligodendrocytes, Schwann cells, and neurons, producing demyelination of the central and peripheral nervous systems and a relentlessly progressive neurodegenerative clinical course. Because GAL3ST1/CST catalyses the terminal, rate-limiting sulfation step that converts galactosylceramide into sulfatide — the immediate biosynthetic precursor of the very lipid that accumulates in MLD — partial pharmacological inhibition of GAL3ST1 has been proposed, at a purely conceptual level, as a substrate-reduction therapy (SRT) strategy: reducing new sulfatide synthesis upstream, rather than attempting to restore degradative ARSA activity, in principle could lessen the biosynthetic load contributing to sulfatide accumulation. This logic mirrors the general substrate-reduction paradigm already used clinically in other lysosomal storage disorders (for example miglustat and eliglustat in Gaucher disease), where partially throttling upstream synthesis is used as an adjunct or alternative to enzyme-directed approaches.

The principal translational motivation for exploring this route computationally, rather than defaulting to conventional enzyme-replacement or gene-therapy approaches, is cost. Currently available disease-modifying treatments for MLD-type lysosomal disorders, including haematopoietic stem-cell gene therapy, are among the most expensive medical interventions in existence, placing them beyond practical reach for the great majority of affected families, particularly in low-

and middle-income settings where MLD is also diagnosed. The central novelty proposed by this project is therefore not a claim of biological efficacy, but a computational, hypothesis-generating exploration of whether inexpensive, dietarily available, and already-consumed plant flavonols — sourced here from Clitoria ternatea (Section 1.4.1) — can be rationally screened, using blind molecular docking against an unresolved and functionally unannotated target, as a low-cost starting point for eventual substrate-reduction research in MLD. No claim is made in this manuscript that GAL3ST1 inhibition would meaningfully alter MLD pathology in vivo, that the flavonols examined cross the blood-brain barrier, or that they constitute an effective or established treatment for MLD; all findings are reported strictly as computational, prediction-only outputs requiring extensive experimental validation before any biological or clinical inference can be drawn (Section 3.10 and Section 7).

1.3 Dietary Flavonols: Quercetin, Kaempferol, and Isorhamnetin

Quercetin, kaempferol, and isorhamnetin are three of the most widely consumed dietary flavonols, sharing the characteristic 3-hydroxyflavone (3-hydroxy-2-phenylchromen-4-one) scaffold and differing from one another only in the substitution pattern of the B-ring. Kaempferol carries a single para-hydroxyl on the B-ring (molecular formula C15H10O6, molecular weight 286.24 g/mol); quercetin carries an additional ortho-hydroxyl, giving a catechol (3',4'-dihydroxyphenyl) B-ring (C15H10O7, 302.24 g/mol); and isorhamnetin is the 3'-O-methylated derivative of quercetin, in which one of the two catechol hydroxyls is capped as a methoxy group (C16H12O7, 316.26 g/mol). This close structural relationship — kaempferol and quercetin differing by a single hydroxyl, and quercetin and isorhamnetin differing by a single methylation — makes the trio a natural, internally controlled series for structure-activity comparison, since any differences observed in docking behaviour, physicochemical profile, or toxicity prediction can be attributed with reasonable confidence to that single, well-defined chemical modification rather than to wholesale differences in molecular architecture.

All three compounds are abundant in the human diet, being present in onions, kale, broccoli, apples, berries, tea, and numerous other fruits and vegetables, and all three have been the subject of extensive pharmacological characterisation, with reported antioxidant, anti-inflammatory, antimicrobial, anti-adipogenic, and anticancer activities described across cell-based and animal models. Their broad-spectrum bioactivity, oral availability from common foods, and long history of human dietary exposure make them attractive starting points for exploring modulation of under-studied enzymatic targets such as GAL3ST1, since a favourable pre-existing safety record (albeit for their general dietary use rather than for any specific pharmacological indication) lowers the practical barrier to eventual experimental follow-up relative to a fully synthetic, previously uncharacterised chemical entity.

1.4.1 Clitoria ternatea (Butterfly Pea Flower): Botanical and Phytochemical Profile

Clitoria ternatea Linn. (family Fabaceae; commonly known as butterfly pea flower, Aparajita, or blue pea) is a perennial, twining legume native to tropical South and Southeast Asia and now cultivated widely across the Indian subcontinent as an ornamental, fodder, and traditional-medicine plant. Botanically, it is characterised by trifoliate pinnate leaves, a climbing herbaceous habit, and a distinctive deep-blue to violet, occasionally white, papilionaceous flower whose colour derives from a dense complement of anthocyanins (principally ternatins and delphinidin-based glycosides). In Ayurvedic and other South Asian traditional-medicine systems, the flower, root, and seed have been used as a memory-enhancing (medhya), anxiolytic, nootropic, anti-inflammatory, diuretic, and antipyretic agent, and aqueous flower extracts remain in widespread culinary use as a natural blue colourant.

Phytochemically, Clitoria ternatea flowers and leaves are a documented source of a broad flavonoid and polyphenolic pool, including the 3-hydroxyflavone-scaffold flavonols examined in the present study, alongside anthocyanins, flavone glycosides, triterpenoids, and phenolic acids. It is this natural co-occurrence of quercetin-, kaempferol-, and isorhamnetin-type constituents within a single, dietarily and traditionally consumed plant matrix that provides the

botanical rationale for treating C. ternatea as a candidate whole-plant source of the flavonol series modelled in this manuscript, rather than selecting the three compounds as isolated, unrelated reference chemicals. Selecting a single plant source is also intended to keep the eventual translational question narrow and answerable: whether extract- or fraction-level preparations of a demonstrably safe, widely consumed botanical can be rationally prioritised, on computational grounds, for further work against an under-characterised enzymatic target, rather than proposing any individual purified compound as a stand-alone drug candidate at this stage.



Figure 1.1: Morphological Features of *Clitoria Ternatea* (Blue Pea): (A) Detailed View of the Vibrant Blue Blossom, and (B) the Whole Plant in a Potted Setup During Its Growth Phase.

1.5 Computational Drug Discovery and the Rationale for Blind Docking

Structure-based virtual screening ordinarily begins from a defined binding site: either a co-crystallised ligand pose, a catalytic residue triad identified by sequence homology, or a site nominated by mutagenesis data. GAL3ST1 offers none of these anchors. It has never been crystallised, it belongs to a sulfotransferase sub-family with limited sequence identity to structurally characterised relatives, and — as emphasised above — the residues that coordinate PAPS or the glycolipid acceptor have not been experimentally mapped for this particular protein. Restricting docking a priori to a manually guessed region of the model would therefore risk anchoring every subsequent result to an incorrect assumption about where the catalytically or pharmacologically relevant surface actually lies.

Blind docking sidesteps this problem by treating the entire protein surface as the search space and allowing an automated cavity-detection algorithm, rather than the investigator, to nominate candidate binding pockets on the basis of surface geometry alone. In this study that role was performed by CB-Dock2, which layers a curvature-based cavity-detection module (CurPocket) beneath the AutoDock Vina docking engine: CurPocket first enumerates and ranks pockets across the entire receptor surface by volume and shape complementarity to a probe ligand, after which Vina performs local, semi-flexible docking within a bounding box centred on each candidate cavity in turn. The output — a ranked list of cavities, each with its own docking score and set of contact residues — is precisely the unbiased, hypothesis-free evidence base that an experimentally uncharacterised target such as GAL3ST1 demands, and it is this rationale, stated explicitly at the outset of the project, that motivated the blind-docking design of the present study rather than a targeted or covalent-docking alternative.

1.6 Aims and Objectives of the Present Study

The present study was designed with the following specific objectives:

- To identify and rank the candidate ligand-binding cavities on a computational model of human GAL3ST1 using cavity-detection-guided blind docking, in the absence of any prior experimental annotation of the active site.

- To dock the three structurally related dietary flavonols quercetin, kaempferol, and isorhamnetin into every identified cavity and to compare their binding scores, poses, and contact-residue fingerprints.
- To characterise the drug-likeness, physicochemical properties, and predicted pharmacokinetic behaviour of the three compounds using SwissADME.
- To profile the predicted acute, organ-level, and pathway-level toxicity of each compound using ProTox-3.0.
- To generate a ligand-similarity-based off-target and polypharmacology profile for each compound using SwissTargetPrediction, and to interpret this profile in the specific context of a target, GAL3ST1, that is not itself represented in the training data of such similarity-based predictors.
- To integrate these four independent lines of computational evidence into a single, internally consistent structure-activity and safety narrative that can inform prioritisation for subsequent in vitro validation.

1.7 Structure of this Manuscript

The remainder of this manuscript is organised as follows. Section 2 reviews the relevant structural-biology, disease-biology, and computational-methodology literature in greater depth than was possible in this introductory section. Section 3 documents the full materials and methods used, including an explicit discussion of why blind docking was the only defensible strategy available for this target and an equally explicit accounting of the study's methodological limitations, stated in advance of the results so that they can be borne in mind while those results are read. Section 4 presents the docking, physicochemical, toxicological, and target-prediction results for each compound in turn before drawing them together comparatively. Section 5 discusses the structural, medicinal-chemistry, safety, and disease-biology implications of those results, explicitly separating well-evidenced conclusions from more speculative interpretation throughout. Sections 6 and 7 close the manuscript with a concise conclusion and a concrete set of recommended next steps, and the References and nine lettered appendices (A through I) that follow provide, respectively, full source citations and complete, unabridged reproductions of every raw dataset summarised in the main text, so that every numerical claim made in this manuscript can be traced back to its original source.

2. Literature Review

2.1 Structural Biology of Sulfotransferases and the PAPS-Binding Fold

Sulfotransferases as a superfamily share a conserved catalytic strategy despite considerable sequence divergence: a bound molecule of PAPS donates its beta-sulfonate group to a nucleophilic hydroxyl (or, in some family members, amine) on the acceptor substrate, typically via an in-line SN2-like transition state stabilised by conserved basic and polar residues that neutralise the developing negative charge on the leaving 3'-phosphoadenosine-5'-phosphate. In the well-studied cytosolic sulfotransferases, this PAPS-binding function is provided by two short sequence motifs — often referred to as the 5'-phosphosulfate-binding (PSB) loop and the 3'-phosphate-binding (PB) motif — which together form a shallow, positively charged groove on the protein surface. GAL3ST1 and its Golgi-resident carbohydrate sulfotransferase relatives are evolutionarily and structurally distinct from the cytosolic sulfotransferases, but they are nonetheless expected, on general biochemical grounds, to require an analogous anion-binding surface to accommodate the doubly phosphorylated PAPS co-substrate. The large, singly dominant cavity identified in the present study (Cavity 1; see Section 4) is lined by a disproportionate number of basic and polar residues — lysine, histidine, arginine, threonine, and serine — relative to the smaller, more hydrophobic secondary cavities, a pattern that is at least consistent with, though not proof of, a PAPS/acceptor-binding role for this cavity; this inference is treated as a working hypothesis rather than an established fact throughout this manuscript, precisely because no experimental mutagenesis or co-crystal data exist to confirm it for GAL3ST1 specifically.

2.2 GAL3ST1 as an Emerging Oncology and Metabolic Target

The oncology literature on GAL3ST1 has concentrated overwhelmingly on renal cell carcinoma, where hypoxic tumour regions upregulate the enzyme through a

HIF1-dependent transcriptional programme, and the resulting sulfatide accumulation on the tumour cell surface has been mechanistically tied to enhanced platelet binding and consequent immune evasion, a phenomenon that offers a biologically plausible route by which pharmacological inhibition of GAL3ST1 could be therapeutically beneficial rather than merely correlative. Corroborating clinical-pathology data associate higher GAL3ST1/Gal3ST1 expression with adverse prognosis in renal cell carcinoma cohorts, and elevated GAL3ST1 transcript levels have separately been reported in ovarian epithelial carcinoma. A smaller but mechanistically important metabolic literature connects common genetic variation at the GAL3ST1 locus to exercise-modified insulin resistance and to type 2 diabetes risk, and rodent injury models show that hepatic Gal3st1 expression, and the circulating sulfatide pool it sustains, respond dynamically to acute kidney injury. No published study, to the authors' knowledge at the time of writing, has reported a validated small-molecule inhibitor of GAL3ST1, structure-based or otherwise; the present work is accordingly framed as an early, hypothesis-generating computational contribution to what is, at present, essentially an open drug-discovery problem.

2.3 Reported Bioactivities of Quercetin, Kaempferol, and Isorhamnetin

Quercetin is among the most extensively studied dietary polyphenols, with a large pharmacological literature documenting antioxidant, anti-inflammatory, antiviral, antihypertensive, and antiproliferative activity across in vitro and rodent models, alongside well-recognised limitations in oral bioavailability driven by extensive first-pass glucuronidation and sulfation. Kaempferol shares much of quercetin's antioxidant and anti-inflammatory activity profile while generally showing a somewhat more favourable pharmacokinetic and safety signature in preclinical studies, and has been separately investigated for anticancer, neuroprotective, and cardioprotective activity. Isorhamnetin, quercetin's 3'-O-methyl metabolite, is both a direct dietary constituent (notably of sea buckthorn) and a phase II metabolite of quercetin itself, and has attracted particular interest for anti-inflammatory, antitumour, and anti-obesity activity, along with the pharmacokinetic advantage that O-methylation typically confers relative to the parent catechol, namely reduced susceptibility to catechol-O-methyltransferase-mediated and oxidative metabolism. None of the existing pharmacological literature on these three flavonols specifically addresses GAL3ST1 or sulfatide metabolism, which is the gap the present computational study is intended to begin addressing.

2.4 Computational Approaches to Studying Under-Characterised Targets

A growing methodological literature has established cavity-detection-guided blind docking as a pragmatic strategy for structure-based work on targets that lack a validated binding-site annotation, a category that includes many enzymes newly implicated in disease through transcriptomic or genetic association studies rather than through classical biochemical characterisation. Tools in this space — of which CB-Dock and its successor CB-Dock2 are among the most widely used, having together processed well over one hundred thousand blind-docking submissions from the research community — combine automated cavity enumeration with conventional physics-based docking scoring, and have been applied across therapeutic areas ranging from antiviral drug repurposing to natural-product target deconvolution. The approach necessarily trades the precision of a validated, literature-anchored active site for breadth of coverage across the entire receptor surface, and its outputs are consequently best interpreted as hypothesis-generating rather than confirmatory, a framing adopted throughout the Results and Discussion sections of this manuscript.

2.5 The Renal Cell Carcinoma Therapeutic Landscape and the Case for New Targets

Clear-cell renal cell carcinoma, the tumour context in which GAL3ST1 biology has been most directly studied (Section 2.2), is currently managed clinically with a combination of anti-angiogenic tyrosine kinase inhibitors, mammalian target of rapamycin (mTOR) inhibitors, and, increasingly, immune checkpoint inhibitors used alone or in combination regimens. Despite meaningful gains in progression-free survival with these agents over the past two decades, acquired resistance remains near-universal, and durable complete responses remain uncommon outside the

subset of patients who respond deeply to checkpoint blockade. The HIF1-GAL3ST1-sulfatide axis described in Section 2.2 is of particular interest in this context precisely because it operates through a distinct biological mechanism — enhanced tumour cell-platelet binding and consequent immune evasion — that is orthogonal to the angiogenic and immune-checkpoint pathways already targeted by existing therapy, raising the possibility, still entirely hypothetical at the present stage of evidence, that a GAL3ST1-directed agent could be developed as an adjunct to, rather than a replacement for, existing renal cell carcinoma regimens. This unmet clinical need for mechanistically distinct combination partners is part of the broader rationale motivating exploratory computational work of the kind reported in this manuscript, even though the present study stops well short of any claim of antitumour efficacy for the three flavonols examined.

2.6 Natural Products and Dietary Phytochemicals in Contemporary Drug Discovery

Natural products and their derivatives continue to make an outsized contribution to the pharmacopoeia relative to their representation in typical synthetic screening libraries, a pattern usually attributed to the evolutionary pre-selection of natural product chemical space for productive interactions with biological macromolecules. Dietary flavonoids occupy a particular niche within this broader natural-product landscape: unlike many plant secondary metabolites, they are consumed daily by essentially the entire human population at gram-scale quantities across a lifetime, generating an unusually large, if uncontrolled, real-world human exposure and safety dataset that is simply unavailable for most synthetic drug candidates at the point of first computational or in vitro characterisation. This pre-existing exposure history does not substitute for formal toxicological and pharmacokinetic characterisation of any given flavonoid for a specific, novel pharmacological indication such as GAL3ST1 inhibition, but it does meaningfully lower the practical and regulatory barrier to eventual clinical translation relative to a de novo synthetic chemical entity, and it is one of the considerations, alongside the structural and mechanistic rationale developed in Sections 1.4 and 1.5, that motivated the selection of quercetin, kaempferol, and isorhamnetin as the ligand set for this study.

2.7 In Silico ADMET and Toxicity Prediction Platforms

Physicochemical filtering and computational toxicology have become standard early triage steps in both academic and industrial small-molecule discovery pipelines. SwissADME provides rapid, consensus-based estimates of lipophilicity, aqueous solubility, gastrointestinal absorption, blood-brain barrier permeation, and cytochrome P450 and P-glycoprotein interaction liability, alongside classical rule-based drug-likeness filters (Lipinski's rule of five, and the Ghose, Veber, Egan, and Muegge filters) and structural-alert screens for pan-assay interference compounds (PAINS) and other undesirable substructures (the Brenk filter). ProTox-3.0, the current iteration of the ProTox toxicity-prediction webserver, extends earlier versions by combining molecular-similarity-based read-across with machine-learning classifiers trained on curated toxicological datasets to predict acute oral lethal dose (LD50) and a battery of organ-toxicity, genotoxicity, and Tox21 nuclear-receptor and stress-response pathway endpoints. SwissTargetPrediction, finally, estimates the most probable macromolecular targets of a small molecule by combining two-dimensional and three-dimensional similarity to a curated library of molecules with experimentally annotated targets drawn chiefly from ChEMBL; a structural corollary of this design, of direct relevance to the present study, is that a target absent from that curated library — as GAL3ST1 currently is, having no annotated small-molecule ligands in ChEMBL — cannot be returned as a hit regardless of a compound's true affinity for it, a limitation discussed explicitly in Section 5.5.

DECLARATIONS

Funding

This computational study did not receive any dedicated external funding. Standard institutional or personal computing resources were used to run all analyses described herein.

Conflicts of Interest

The author(s) declare no competing financial or non-financial interests related to the subject matter of this manuscript.

Author Contributions

Conceptualisation, computational analysis, data curation, and manuscript preparation: [to be completed by corresponding author]. All authors read and approved the final manuscript.

Data Availability Statement

The receptor coordinate file, ligand structures, CB-Dock2 cavity-detection tables, Swiss ADME outputs, ProTox-3.0 reports, and Swiss Target Prediction reports underlying this manuscript are available from the corresponding author on reasonable request. No new experimental data were generated.

Ethical Approval

This study is purely computational (in silico) in nature and did not involve human participants, human data, or animal subjects. Ethical approval was therefore not required.

Acknowledgements

The authors acknowledge the developers of CB-Dock2, Auto Dock Vina, Swiss ADME, Swiss Target Prediction, and ProTox-3.0 for making these platforms freely available to the research community.

REFERENCES

1. Appendix A: Full CB-Dock2 Cavity Tables 84 A.1 Quercetin 84 A.2 Kaempferol 85 A.3 Isorhamnetin (attributed; see Section 3.2 and Section 5.7) 87 Appendix B: ProTox-3.0 Toxicity Target Pharmacophore Data 89 B.1 Kaempferol 89 B.2 Isorhamnetin 89 B.3 Quercetin 89 Appendix C: Extended Swiss Target Prediction Tables 91 C.1 Kaempferol - extended target list (top 45) 91 C.2 Isorhamnetin - extended target list (top 45) 93 C.3 Quercetin - extended target list (top 45) 95 Appendix D: Full Single-Compound ProTox-3.0 Reports 98 D.1 Kaempferol 98 D.2 Isorhamnetin 100 D.3 Quercetin 102 Appendix E: Ligand Identifiers and Structures 105 Appendix F: Glossary of Technical Terms 106 Appendix G: Data Provenance and Reporting Checklist 108 Appendix H: Statement on the Use of Computational and AI Tools 110 Appendix I: Outline Protocol for Experimental Validation 111 I.1 Recombinant Enzyme Activity Assay 111 I.2 Catechol-Reactivity Counter-Screen (Quercetin Arm Only) 111 I.3 Off-Target Counter-Screens 111 I.4 Cell-Based Phenotypic Follow-Up 112 Note on referencing: the entries below combine (i) sources whose bibliographic details were directly verified for this manuscript, particularly those concerning GAL3ST1 biology and the specific software tools used (CB-Dock/CB-Dock2, AutoDock Vina, SwissADME, SwissTargetPrediction, ProTox), and (ii) well-established landmark methodological papers (the Lipinski, Ghose, Veber, Egan, Muegge, PAINS, and Brenk filters, and general flavonoid pharmacology reviews) cited from standing scientific knowledge. Readers intending to submit a derivative of this manuscript for formal publication should independently re-verify exact volume, issue, and page-number details against the original sources before submission.
2. Liu Y, Yang X, Gan J, Chen S, Xiao Z-X, Cao Y. CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. Nucleic Acids Research. 2022;50 (W1):W159-W164.
3. Liu Y, Grimm M, Dai W, Hou M, Xiao Z-X, Cao Y. CB-Dock: a web server for cavity detection-guided protein-ligand blind docking. Acta Pharmacologica Sinica. 2020;41:138-144.
4. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry. 2010;31 (2):455-461.
5. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports. 2017;7:42717.

6. Daina A, Michielin O, Zoete V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. Nucleic Acids Research. 2019;47 (W1):W357-W364.
7. Banerjee P, Eckert AO, Schrey AK, Preissner R. ProTox-II: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Research. 2018;46 (W1):W257-W263.
8. Charite Universitatsmedizin Berlin. ProTox-3.0: Prediction of Toxicity of Chemicals [Internet webserver]. Available at: https://tox.charite.de/protox3.
9. Honke K, Hirahara Y, Dupree J, et al. Paranodal junction formation and spermatogenesis require sulfoglycolipids. Proceedings of the National Academy of Sciences USA. 2002;99 (7):4227-4232.
10. [Author (s) not independently verified in this search]. A Hypoxia-Inducible HIF1-GAL3ST1-Sulfatide Axis Enhances ccRCC Immune Evasion via Increased Tumor Cell-Platelet Binding. 2019. PMID: 31427440.
11. [Author (s) not independently verified in this search]. The prognostic value of galactosylceramide-sulfotransferase (Gal3ST1) in human renal cell carcinoma. 2021. PMID: 34035403.
12. Zhang X, Nakajima T, Kamijo Y, et al. Acute kidney injury induced by protein-overload nephropathy down-regulates gene expression of hepatic cerebroside sulfotransferase in mice, resulting in reduction of liver and serum sulfatides. Biochemical and Biophysical Research Communications. 2009;390 (4):1382-1388.
13. [Author (s) not independently verified in this search]. A variation in the cerebroside sulfotransferase gene is linked to exercise-modified insulin resistance and to type 2 diabetes. 2009. PMID: 19587831.
14. UniProt Consortium. UniProtKB entry Q99999 (GAL3ST1_HUMAN, Galactosylceramide sulfotransferase). Available at: https://www.uniprot.org/uniprotkb/Q99999.
15. National Center for Biotechnology Information. GAL3ST1 galactose-3-O-sulfotransferase 1 [Homo sapiens], Gene ID: 9514. NCBI Gene database.
16. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 1997/2001;46 (1-3):3-26.
17. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. Journal of Medicinal Chemistry. 2002;45 (12):2615-2623.
18. Egan WJ, Merz KM, Baldwin JJ. Prediction of drug absorption using multivariate statistics. Journal of Medicinal Chemistry. 2000;43 (21):3867-3877.
19. Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. Journal of Combinatorial Chemistry. 1999;1 (1):55-68.
20. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. Journal of Medicinal Chemistry. 2010;53 (7):2719-2740.
21. Brenk R, Schipani A, James D, et al. Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. ChemMedChem. 2008;3 (3):435-444.
22. Panche AN, Diwan AD, Chandra SR. Flavonoids: an overview. Journal of Nutritional Science. 2016;5:e47.
23. D'Andrea G. Quercetin: a flavonol with multifaceted therapeutic applications? Fitoterapia. 2015;106:256-271.
24. Calderon-Montano JM, Burgos-Moron E, Perez-Guerrero C, Lopez-Lazaro M. A review on the dietary flavonoid kaempferol. Mini-Reviews in Medicinal Chemistry. 2011;11 (4):298-344.

25. Gong G, Guan YY, Zhang ZL, et al. Isorhamnetin: a review of pharmacological effects. Biomedicine and Pharmacotherapy. 2020;128:110301.
26. GeneCards Human Gene Database. GAL3ST1 Gene - Galactose-3-O-Sulfotransferase 1. Weizmann Institute of Science.

Appendix A: Full CB-Dock2 Cavity Tables The following tables reproduce, in full and without abridgement, the raw CB-Dock2 cavity-detection and blind-docking output for each of the three flavonols across all five identified cavities on the GAL3ST1 model, as summarised in Sections 4.2 to 4.4.

A.1 Quercetin

.1 Quercetin

Cavity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
1	14326	-3.896, 0.980, -9.535	35 x 35 x 31	-9.6	LYS82, THR83, HIS84, LYS85, THR86, ALA87, SER88, SER89, THR90, ARG112, ASN113, ASP114, TYR117, HIS141, ARG164, GLU171, SER172, HIS175, TYR176, PHE177, VAL180, ASN210, TYR213, GLN252, LYS277, LEU278, ASN279, ALA280, ARG281, ARG282, ASP283, SER284, ARG288, PRO364, TRP365, ILE373, PRO396, GLU397, ILE398, THR411, TRP414
2	1115	-4.497, 4.617, 23.261	21 x 21 x 21	-6.7	GLU58, PRO59, GLU60, ALA61, VAL62, ILE63, ARG64, CYS72, GLN73, PRO74, ARG75, ARG76, ASN77, ASN136, ASN156
3	490	-17.660, -11.575, 10.160	21 x 21 x 21	-7.0	ILE94, ARG97, PHE98, GLN100, LYS101, GLU271, ASP272, VAL273, LEU274, MET341, GLN384, ARG385, HIS386, ALA387, GLN388, LEU389, CYS390, ARG392
4	210	-6.493, 18.397, -5.652	21 x 21 x 21	-7.6	TYR30, SER31, TYR32, ALA33, VAL34, PRO35, PRO36, HIS84, PRO109, ASN110, GLY111, ARG112, ASN113, ASP114, PHE115, ASP116, TYR117, PRO118, SER124, LEU125, VAL126, GLN127, PHE177, VAL180, VAL181, PHE209, ASN210, HIS212, TYR213,

Cavity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
					TRP365, LEU368, TRP409, VAL410, THR411, LYS412, TRP414, LYS415, ILE417, ARG418, TRP423
5	198	-7.391, -7.356, -17.035	21 x 21 x 21	-5.9	LYS85, SER172, PHE174, HIS175, TYR176, PHE177, PRO179, GLY190, LYS192, LYS277, ALA280, ARG281, ARG282, ASP283, SER284, PRO285, VAL286, GLY349, HIS350, ALA351, VAL352, ASP353, ALA355, PRO367, LEU368, GLY369, THR370, LYS371, SER372, LEU374, TYR376

A.2 Kaempferol

Cavity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
1	14326	-3.896, 0.980, -9.535	35 x 35 x 31	-9.6	LYS82, THR83, HIS84, LYS85, THR86, ALA87, SER88, SER89, THR90, ASN113, ASP114, TYR117, HIS141, ARG164, ARG168, SER172, TYR176, PHE177, VAL180, ASN210, TYR213, GLN252, LEU278, ASN279, ALA280, ARG281, PRO364, TRP365, LEU368, ILE373, PRO396, GLU397, ILE398, GLN399, THR411
2	1115	-4.497, 4.617, 23.261	21 x 21 x 21	-6.4	CYS52, PRO54, PRO55, ALA56, LEU57, GLU58, PRO59, GLU60, ALA61, VAL62, ILE63, CYS72, GLN73, PRO74, ARG75, ARG76, ASN77, LEU104, ASN136, PRO154, THR155, ASN156, ILE158
3	490	-17.660, -11.575, 10.160	21 x 21 x 21	-6.7	ASN93, ILE94, PHE96, ARG97, PHE98, GLN100, LYS101, GLU269, GLU271, ASP272, VAL273, LEU274, GLN384,

Cav-ity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
					ARG385, HIS386, ALA387, GLN388, LEU389, CYS390, ARG392, TYR400
4	210	-6.493, 18.397, -5.652	21 x 21 x 21	-7.5	TYR30, SER31, TYR32, ALA33, VAL34, PRO35, PRO36, PRO109, ASN110, GLY111, ARG112, ASN113, ASP114, PHE115, ASP116, TYR117, PRO118, SER124, LEU125, VAL126, PHE177, VAL180, PHE209, ASN210, HIS212, TYR213, TRP365, GLN366, LEU368, TRP409, VAL410, THR411, TRP414, LYS415, ARG418, TRP423
5	198	-7.391, -7.356, -17.035	21 x 21 x 21	-5.9	LYS85, ARG164, ARG168, GLU171, SER172, PHE174, HIS175, TYR176, PHE177, GLY178, PRO179, THR184, LYS192, LYS277, ASN279, ALA280, ARG281, ARG282, ASP283, SER284, PRO285, VAL286, ARG288, ARG342, GLY349, HIS350, ALA351, ASP353, PRO367, LEU368, GLY369, THR370, SER372, LEU374, TYR376

A.3 Isorhamnetin (attributed; see Section 3.2 and Section 5.7)

Cav-ity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
1	14326	-3.896, 0.980, -9.535	35 x 35 x 31	-9.9	VAL34, PRO35, PRO36, LEU37, LYS82, THR83, HIS84, LYS85, THR86, ALA87, SER88, SER89, ARG112, ASN113, ASP114, TYR117, PRO118, THR119, PHE120, HIS141, ARG143, ARG164, SER172, TYR176, PHE177, VAL180, TYR204, PRO206, GLY208, PHE209, ASN210, ALA211, HIS212, TYR213, PHE220, ASP225, ASN226, SER227, LEU278, ASN279, ALA280, ARG281, TRP365, PRO367, LEU368, ILE373, GLU397, ILE398, THR411, TRP414
2	1115	-4.497, 4.617, 23.261	21 x 21 x 21	-6.4	SER51, CYS52, PRO54, PRO55, LEU57, GLU58, PRO59, GLU60, ALA61, VAL62, ILE63, ARG64, GLU71, CYS72, GLN73, PRO74, ARG75, ARG76, ASN77, ASN136, ARG150, PRO154, THR155, ASN156
3	490	-17.660, -11.575, 10.160	21 x 21 x 21	-7.3	ILE94, ARG97, PHE98, GLN100, LYS101, GLU269, GLU271, ASP272, VAL273, LEU274, MET341, ARG385, HIS386, GLN388, LEU389, ARG392, TYR400
4	210	-6.493, 18.397, -5.652	21 x 21 x 21	-7.5	TYR30, SER31, TYR32, ALA33, VAL34, PRO35, PRO36, PRO109, ASN110, GLY111, ARG112, ASN113, ASP114, PHE115, ASP116, TYR117, PRO118, SER124, LEU125, VAL126, PHE177, VAL180, PHE209, ASN210, HIS212, TYR213, TRP365, LEU368, TRP409, VAL410, THR411, TRP414, LYS415, ARG418, TRP423

Cavity	Volume (Å³)	Center (x, y, z)	Grid size (Å)	Vina score (kcal/mol)	Key contact residues
5	198	-7.391, -7.356, -17.035	21 x 21 x 21	-6.0	LYS85, SER172, PHE174, HIS175, TYR176, PHE177, GLY178, PRO179, THR184, LYS192, ASN279, ARG281, ARG282, ASP283, SER284, PRO285, VAL286, ASP353, ALA354, ALA355, PRO367, LEU368, GLY369, THR370, LYS371, SER372, ILE373

Appendix B: ProTox-3.0 Toxicity Target Pharmacophore Data

In addition to the organ-, pathway-, and mechanism-level endpoints tabulated in Table 4.4, ProTox-3.0 reports an average pharmacophore fit and average similarity to known ligands for a small set of curated toxicity targets. These supplementary metrics are reproduced below for completeness.

B.1 Kaempferol

Toxicity target	Avg. pharmacophore fit	Avg. similarity, known ligands
Androgen Receptor	1.04%	83.77%
Amine Oxidase A	26.87%	88.12%
Dopamine Receptor D3	0%	73.11%
Prostaglandin G/H Synthase 1	33.6%	83%

B.2 Isorhamnetin

Toxicity target	Avg. pharmacophore fit	Avg. similarity, known ligands
Androgen Receptor	2.44%	76.55%
Amine Oxidase A	43.36%	82.73%
Estrogen Receptor 2	0.83%	81.08%
Prostaglandin G/H Synthase 1	47.54%	76.79%

B.3 Quercetin

Toxicity target	Avg. pharmacophore fit	Avg. similarity, known ligands
Androgen Receptor	1.69%	79.43%
Amine Oxidase A	36.22%	90.1%

Toxicity target	Avg. pharmacophore fit	Avg. similarity, known ligands
Estrogen Receptor 2	0.39%	72.97%
Prostaglandin G/H Synthase 1	42.27%	83.51%

Prediction accuracy and reference-compound similarity for the underlying LD50 model, reported directly by ProTox-3.0, were: kaempferol, 82.46% average similarity, 70.97% prediction accuracy; isorhamnetin, 87.48% average similarity, 70.97% prediction accuracy; quercetin, 100% average similarity, 100% prediction accuracy. As discussed in Section 4.8 and Section 5.4, this places materially greater direct evidentiary weight behind the quercetin acute-toxicity classification than behind the kaempferol or isorhamnetin classifications.

Appendix C: Extended Swiss Target Prediction Tables

The tables below extend Tables 4.5 through 4.7 to the top predicted targets retrieved for each compound, provided here for readers wishing to examine the full poly pharmacology landscape beyond the top-ten summary presented in the main Results.

C.1 Kaempferol - extended target list (top 45)

Predicted target	Target class	Probability
ABCG2	Primary active transporter	1.00
TRPC5	Voltage-gated ion channel	1.00
IP6K2	Transferase	1.00
HDAC6	Eraser	1.00
CISD1	Unclassified protein	1.00
NOX4	Oxidoreductase	1.00
IPMK	Transferase	1.00
CYP1B1	Cytochrome P450	1.00
RPS6KA3	Kinase	1.00
XDH	Oxidoreductase	1.00
CA7	Lyase	1.00
HSD17B2	Oxidoreductase	1.00
FLT3	Kinase	1.00
AHR	Transcription factor	1.00
ABCC1	Primary active transporter	1.00

Predicted target	Target class	Probability
DPP4	Protease	1.00
TYR	Oxidoreductase	1.00
HSD17B1	Oxidoreductase	1.00
ESRRA	Nuclear receptor	1.00
SLC2A1	Electrochemical transporter	1.00
ALOX5	Oxidoreductase	1.00
ABCB1	Primary active transporter	1.00
YES1	Kinase	1.00
MPO	Oxidoreductase	1.00
CA2	Lyase	1.00
CA12	Lyase	1.00
ACHE	Hydrolase	0.988
CA4	Lyase	0.988
MAOA	Oxidoreductase	0.988
PPARG	Nuclear receptor	0.984
SETD7	Writer	0.983
SLC16A1	Electrochemical transporter	0.983
APP	Membrane receptor	0.983
PTPRS	Phosphatase	0.983
ALOX15	Oxidoreductase	0.981
CSNK2A1	Kinase	0.979
ALOX12	Enzyme	0.979
PKM	Enzyme	0.979
SLCO2B1	Electrochemical transporter	0.979
SIRT6	Eraser	0.978
METTL3/METTL14	Transferase	0.978
GLO1	Lyase	0.978

Predicted target	Target class	Probability
BACE1	Protease	0.978
GSK3B	Kinase	0.978
SYK	Kinase	0.978

C.2 Isorhamnetin - extended target list (top 45)

Predicted target	Target class	Probability
IP6K2	Transferase	1.00
IPMK	Transferase	1.00
CYP1B1	Cytochrome P450	1.00
XDH	Oxidoreductase	1.00
CA7	Lyase	1.00
PPARG	Nuclear receptor	1.00
CA4	Lyase	1.00
CA2	Lyase	1.00
CA12	Lyase	1.00
ELAVL3	Unclassified protein	0.988
ACHE	Hydrolase	0.988
CA4	Lyase	0.988
MAOA	Oxidoreductase	0.988
MAOB	Oxidoreductase	0.967
BCHE	Hydrolase	0.967
SNCA	Unclassified protein	0.966
ADORA3	GPCR (Family A)	0.962
P4HB	Isomerase	0.962
ELAVL3	Unclassified protein	0.961
RPS6KA1	Kinase	0.961
RPS6KA2	Kinase	0.961

Predicted target	Target class	Probability
CDK5/CDK5R1	Kinase	0.961
CDK6	Kinase	0.961
CDK2	Kinase	0.961
CDK1/cyclin B	Kinase	0.961
FBP1	Phosphatase	0.959
CA9	Lyase	0.959
CA5A	Lyase	0.959
CA6	Lyase	0.959
CA3	Lyase	0.959
CA1	Lyase	0.959
SRC	Kinase	0.957
NEK6	Kinase	0.957
PKN1	Kinase	0.957
PTK2	Kinase	0.957
NEK2	Kinase	0.957
AKT1	Kinase	0.957
NUAK1	Kinase	0.957
MKNK2	Kinase	0.953
TNKS2	Transferase	0.953
TTR	Secreted protein	0.953
TNKS	Transferase	0.953
AKR1B10	Oxidoreductase	0.953
PIK3CA	Transferase	0.952
CTNNB1	Unclassified protein	0.952
EGFR	Kinase	0.952

C.3 Quercetin - extended target list (top 45)

Predicted target	Target class	Probability
ABCG2	Primary active transporter	1.00
ALK	Kinase	1.00
CA14	Lyase	1.00
TRPC5	Voltage-gated ion channel	1.00
IP6K2	Transferase	1.00
HDAC6	Eraser	1.00
SPHK1	Enzyme	1.00
NOX4	Oxidoreductase	1.00
NEK6	Kinase	1.00
AURKB	Kinase	1.00
DPP4	Protease	1.00
ALOX5	Oxidoreductase	1.00
ABCB1	Primary active transporter	1.00
YES1	Kinase	1.00
MPO	Oxidoreductase	1.00
SNCA	Unclassified protein	0.966
CTNNB1	Unclassified protein	0.965
CSNK2A1	Kinase	0.965
ACHE	Hydrolase	0.965
ALOX12	Enzyme	0.965
ALOX15	Oxidoreductase	0.965
PKM	Enzyme	0.965
SLCO2B1	Electrochemical transporter	0.965
CDK2	Kinase	0.963
SIRT6	Eraser	0.963
METTL3/METTL14	Transferase	0.963
GLO1	Lyase	0.963

Predicted target	Target class	Probability
BACE1	Protease	0.963
GSK3B	Kinase	0.963
SYK	Kinase	0.963
PSMB5	Protease	0.963
MMP9	Protease	0.963
CYP19A1 (Aromatase)	Cytochrome P450	0.963
MMP2	Protease	0.963
KDR	Kinase	0.963
SRC	Kinase	0.963
MET	Kinase	0.963
CA9	Lyase	0.961
CA5A	Lyase	0.961
CA6	Lyase	0.961
CA3	Lyase	0.961
CA1	Lyase	0.961
AURKB	Kinase	0.960
PIM1	Kinase	0.960
MKNK2	Kinase	0.956

Appendix D: Full Single-Compound ProTox-3.0 Reports

For completeness and ease of independent cross-checking, the full ProTox-3.0 classification report for each compound is reproduced individually below (rather than in the side-by-side comparative format of Table 4.4), together with the physicochemical descriptors ProTox itself computed for each input structure.

D.1 Kaempferol

Property	Value
Molecular weight (g/mol)	286.24
H-bond acceptors	5
H-bond donors	4
Number of atoms	21

Property	**Value**
Number of bonds	23
Rotatable bonds	1
Molecular refractivity	76.01
TPSA (Å²)	111.13
logP (octanol/water)	2.28
Predicted LD50 (mg/kg)	3919
Predicted toxicity class	Class 5 (may be harmful if swallowed)
Average similarity to training set	82.46%
Reported prediction accuracy	70.97%
Endpoint	**Prediction (probability)**
Hepatotoxicity	Inactive (0.68)
Neurotoxicity	Inactive (0.89)
Nephrotoxicity	Active (0.62)
Respiratory toxicity	Active (0.83)
Cardiotoxicity	Inactive (0.91)
Carcinogenicity	Inactive (0.72)
Immunotoxicity	Inactive (0.96)
Mutagenicity	Inactive (0.52)
Cytotoxicity	Inactive (0.98)
BBB-barrier	Active (0.57)
Ecotoxicity	Inactive (0.51)
Clinical toxicity	Inactive (0.54)
Nutritional toxicity	Active (0.66)
AhR	Active (1.00)
Androgen Receptor	Inactive (0.99)
AR-LBD	Inactive (0.99)
Aromatase	Active (0.96)

Property	Value
ERα	Active (1.00)
ER-LBD	Active (0.95)
PPAR-γ	Inactive (0.95)
NRF2/ARE	Inactive (0.99)
HSE	Inactive (0.99)
MMP	Active (1.00)
p53	Inactive (0.92)
ATAD5	Inactive (0.92)
CYP1A2	Active (0.93)
CYP2C19	Active (0.65)
CYP2C9	Active (0.66)
CYP2D6	Active (0.62)
CYP3A4	Inactive (0.65)
CYP2E1	Inactive (1.00)

D.2 Isorhamnetin

Property	Value
Molecular weight (g/mol)	316.26
H-bond acceptors	6
H-bond donors	4
Number of atoms	23
Number of bonds	25
Rotatable bonds	2
Molecular refractivity	82.50
TPSA (Å²)	120.36
logP (octanol/water)	2.29
Predicted LD50 (mg/kg)	5000
Predicted toxicity class	Class 5 (may be harmful if swallowed)

Property	Value
Average similarity to training set	87.48%
Reported prediction accuracy	70.97%
Endpoint	**Prediction (probability)**
Hepatotoxicity	Inactive (0.72)
Neurotoxicity	Inactive (0.88)
Nephrotoxicity	Active (0.64)
Respiratory toxicity	Active (0.85)
Cardiotoxicity	Active (0.82)
Carcinogenicity	Inactive (0.68)
Immunotoxicity	Active (0.58)
Mutagenicity	Inactive (0.94)
Cytotoxicity	Inactive (0.95)
BBB-barrier	Active (0.52)
Ecotoxicity	Inactive (0.55)
Clinical toxicity	Inactive (0.50)
Nutritional toxicity	Active (0.57)
AhR	Active (0.97)
Androgen Receptor	Inactive (1.00)
AR-LBD	Inactive (0.99)
Aromatase	Active (0.58)
ERα	Active (0.88)
ER-LBD	Active (0.89)
PPAR-γ	Inactive (0.95)
NRF2/ARE	Inactive (0.96)
HSE	Inactive (0.96)
MMP	Active (0.92)
p53	Inactive (0.86)

Property	Value
ATAD5	Active (0.65)
CYP1A2	Active (0.91)
CYP2C19	Active (0.83)
CYP2C9	Active (0.63)
CYP2D6	Inactive (0.56)
CYP3A4	Inactive (0.94)
CYP2E1	Inactive (0.99)

D.3 Quercetin

Property	Value
Molecular weight (g/mol)	302.24
H-bond acceptors	6
H-bond donors	5
Number of atoms	22
Number of bonds	24
Rotatable bonds	1
Molecular refractivity	78.04
TPSA (Å²)	131.36
logP (octanol/water)	1.99
Predicted LD50 (mg/kg)	159
Predicted toxicity class	Class 3 (toxic if swallowed)
Average similarity to training set	100%
Reported prediction accuracy	100%
Endpoint	**Prediction (probability)**
Hepatotoxicity	Inactive (0.69)
Neurotoxicity	Inactive (0.89)
Nephrotoxicity	Active (0.62)
Respiratory toxicity	Active (0.83)

Property	Value
Cardiotoxicity	Inactive (0.99)
Carcinogenicity	Active (0.68)
Immunotoxicity	Inactive (0.87)
Mutagenicity	Active (0.51)
Cytotoxicity	Inactive (0.99)
BBB-barrier	Active (0.53)
Ecotoxicity	Inactive (0.53)
Clinical toxicity	Inactive (0.53)
Nutritional toxicity	Active (0.63)
AhR	Active (0.91)
Androgen Receptor	Inactive (0.99)
AR-LBD	Inactive (0.97)
Aromatase	Inactive (0.91)
ERα	Active (0.87)
ER-LBD	Active (0.95)
PPAR-γ	Inactive (0.98)
NRF2/ARE	Inactive (0.99)
HSE	Inactive (0.99)
MMP	Active (1.00)
p53	Inactive (0.97)
ATAD5	Inactive (0.99)
CYP1A2	Active (1.00)
CYP2C19	Active (0.77)
CYP2C9	Active (0.99)
CYP2D6	Inactive (0.85)
CYP3A4	Inactive (0.79)
CYP2E1	Inactive (1.00)

Property	Value

Appendix E: Ligand Identifiers and Structures

Table E.1 summarises the core chemical identifiers of the three flavonols examined in this study, consistent with their PubChem reference entries and cross-validated against the molecular formulae and weights independently reported by Swiss ADME (Section 4.6).

Compound	Formula	MW (g/mol)	Canonical SMILES
Quercetin	C15H10O7	302.24	O=c1c(O)c(-c2ccc(O)c(O)c2)oc2cc(O)cc(O)c12
Kaempferol	C15H10O6	286.24	O=c1c(O)c(-c2ccc(O)cc2)oc2cc(O)cc(O)c12
Isorhamnetin	C16H12O7	316.26	COc1cc(-c2oc3cc(O)cc(O)c3c(=O)c2O)ccc1O

Two-dimensional structures and radar-plot lipophilicity/polarity/size/insaturation/flexibility ("BOILED-Egg"-adjacent) summaries for all three compounds are reproduced as Figures 4.10 to 4.12 in Section 4.6 and 4.7, sourced directly from the SwissADME platform output generated for this study.

Appendix F: Glossary of Technical Terms

Term	Definition
ADMET	Umbrella term for a compound's Absorption, Distribution, Metabolism, Excretion, and Toxicity properties, together the pharmacokinetic and safety profile that determines whether a chemically active molecule can function as a viable oral drug.
Blind docking	A molecular docking strategy in which the entire surface of a receptor is searched for viable ligand-binding cavities, rather than restricting the search to a single, pre-selected active site.
Cavity / pocket	A concave, solvent-accessible depression on a protein surface with the size and chemical character potentially able to accommodate a small-molecule ligand.
Cur Pocket	The curvature-based geometric cavity-detection algorithm underlying the CB-Dock and CB-Dock2 blind-docking platforms.
Consensus Log Po/w	The arithmetic mean of several independent computational estimates of a compound's octanol-water partition coefficient, used by Swiss ADME as a single summary lipophilicity value.
Docking score (Vina score)	The predicted binding free energy, in kilocalories per mole, returned by the Auto Dock Vina scoring function for a given ligand pose within a docking search space; more negative values indicate a more energetically favourable predicted interaction.
Drug-likeness filter	A rule-based heuristic (e.g., Lipinski's rule of five, or the Ghose, Veber,

Term	Definition
	Egan, or Muegge filters) that flags molecules whose physicochemical properties fall outside ranges empirically associated with successful oral drugs.
GHS toxicity class	A six-tier acute-toxicity classification scheme, aligned with the United Nations Globally Harmonised System of Classification and Labelling of Chemicals, ranging from Class 1 (fatal if swallowed) to Class 6 (non-toxic).
LD50	The median lethal dose: the amount of a substance, usually expressed in milligrams per kilogram of body weight, required to kill fifty percent of a test population under specified conditions; used here purely as a computationally predicted, literature-calibrated hazard estimate.
PAINS	Pan-Assay Interference Compounds: substructural motifs statistically associated with non-specific, assay-independent activity in high-throughput screening, used to flag compounds warranting additional counter-screening.
PAPS	3'-Phosphoadenosine-5'-phosphosulfate, the universal biological sulfate-group donor used by all sulfotransferase enzymes, including GAL3ST1.
Sulfatide	3-O-sulfogalactosylceramide, the sulfated glycosphingolipid product of the GAL3ST1 reaction and a major structural lipid of the myelin sheath.
Target class	The broad functional family (e.g., kinase, lyase, oxidoreductase, nuclear receptor) to which a predicted or confirmed protein target belongs, as classified by platforms such as Swiss Target Prediction.
TPSA	Topological Polar Surface Area: the summed surface contribution of polar atoms (principally oxygen and nitrogen and their attached hydrogens) in a molecule, a strong computational correlate of membrane permeability and blood-brain barrier penetration.
Type II transmembrane protein	A membrane protein oriented with its N-terminus on the cytosolic side of the membrane and its larger functional domain on the opposite (in this case, Golgi-luminal) side, the topology predicted for GAL3ST1.

Appendix G: Data Provenance and Reporting Checklist

For transparency, Table G.1 maps every major figure and table in this manuscript to the specific source dataset or platform output from which it was derived, so that the chain of evidence from raw computational output to reported finding can be independently traced and re-verified.

Manuscript item	Source dataset / platform
Tables 4.1-4.3, A.1-A.3	CB-Dock2 Cur Pockets cavity-detection output files (per-

Manuscript item	Source dataset / platform
	compound)
Figures 4.1-4.3	Derived visualisations generated from the CB-Dock2 output and the exported quercetin-Cavity 1 complex coordinate file
Table 4.4, Appendix B, Appendix D	ProTox-3.0 oral toxicity prediction reports (per-compound PDF exports)
Figure 4.6	Derived visualisation of ProTox-3.0 LD50/class outputs
Figures 4.4-4.5	Derived visualisations of Swiss ADME physicochemical descriptors
Figures 4.10-4.12	Direct Swiss ADME platform output panels (source screen-shots)
Tables 4.5-4.7, Appendix C	Swiss Target Prediction ranked target-list reports (per-compound)
Figures 4.7-4.9	Direct Swiss Target Prediction target-class pie-chart output (source screenshots)
Section 4.10	Boltz affinity summary dataset (noted as unrelated to the flavonol series; excluded from quantitative synthesis)

Appendix H: Statement on the Use of Computational and AI Tools

In keeping with current norms of transparency in computational and AI-assisted research reporting, the authors note explicitly that this manuscript integrates output from several third-party computational web-servers and prediction platforms (CB-Dock2, Auto Dock Vina, Swiss ADME, ProTox-3.0, and Swiss Target Prediction), each cited in the References section, and that no experimental wet-laboratory data were generated, reviewed, or claimed anywhere in this manuscript. General-purpose language-model assistance was used in the drafting and organisation of this manuscript's narrative text from the underlying computational datasets described above; all quantitative results, tables, and figures were derived directly and traceably from those datasets as documented in Appendix G, and readers are encouraged to independently verify any finding of particular interest against the original platform outputs before relying on it for further research or translational decision-making.

Appendix I: Outline Protocol for Experimental Validation

For readers wishing to move directly from this computational manuscript toward bench validation, the following outline protocol summarises a minimal, tractable first experimental workstream consistent with the priorities identified in Section 7.

I.1 Recombinant Enzyme Activity Assay

Human GAL3ST1 catalytic domain (or full-length enzyme reconstituted in a Golgi-mimetic membrane preparation) would be expressed recombinantly and incubated with galactosylceramide acceptor and radiolabelled or fluorescently tagged PAPS donor in the presence of a concentration range of each flavonol (typically 0.1 to 100 micromolar, spanning the low end of dietary-relevant exposure through to concentrations more typical of pharmacological in vitro characterisation). Residual sulfotransferase activity would be quantified by thin-layer chromatography, liquid chromatography-mass spectrometry, or a suitable coupled colorimetric assay for the co-product adenosine 3',5'-bisphosphate, and a half-maximal inhibitory concentration (IC50) would be derived for each compound from the resulting concentration-response curve.

I.2 Catechol-Reactivity Counter-Screen (Quercetin Arm Only)

In parallel with the primary activity assay, quercetin's inhibitory activity would be re-tested in the presence of a reducing agent (for example, dithiothreitol or tris(2-carboxyethyl)phosphine) and, separately, in the presence of a metal chelator (for example, EDTA), to test whether inhibition is attenuated under conditions that suppress catechol auto-oxidation or non-specific metal-mediated reactivity, directly addressing the PAINS/Brenk-flagged liability identified in Sections 4.6 and 5.3.

I.3 Off-Target Counter-Screens

Given the Swiss Target Prediction findings summarised in Section 4.9, parallel counter-screens against at least one representative carbonic anhydrase isoform (for example CA2, using a standard esterase or CO2 hydration assay) and against ABCG2-mediated efflux (for example, a calcein-AM or Hoechst 33342 efflux assay in an ABCG2-overexpressing cell line) would help establish whether any downstream cell-based phenotype is attributable specifically to GAL3ST1 modulation rather than to these better-characterised off-target activities.

I.4 Cell-Based Phenotypic Follow-Up

Should recombinant-enzyme inhibition be confirmed, a renal cell carcinoma cell line with documented HIF1-GAL3ST1-sulfatide axis activity would be treated with each active flavonol at non-cytotoxic concentrations, with cell-surface sulfatide quantified by flow cytometry or thin-layer chromatography of extracted lipids, and tumour cell-platelet binding assessed by a standard in vitro co-incubation and flow cytometric or microscopy-based adhesion assay, directly testing the mechanistic hypothesis outlined in Sections 2.2 and 5.7.

HOW TO CITE: *R. Elakkiya*, R. Praveenkumar, R. Riyaskhan, I. Vinnarivu, A Focused Computational Study of *Clitoria Ternatea* Constituents Against GAL3ST1 for a Substrate-Reduction Strategy in MLD., Int. J. of Pharm. Sci., 2026, Vol 4, Issue 11, 4794-4853. https://doi.org/10.5281/zenodo.24548653