

**Journal of Biomolecular Structure and Dynamics**

**ISSN: (Print) (Online) Journal homepage:** [**https://www.tandfonline.com/loi/tbsd20**](https://www.tandfonline.com/loi/tbsd20)

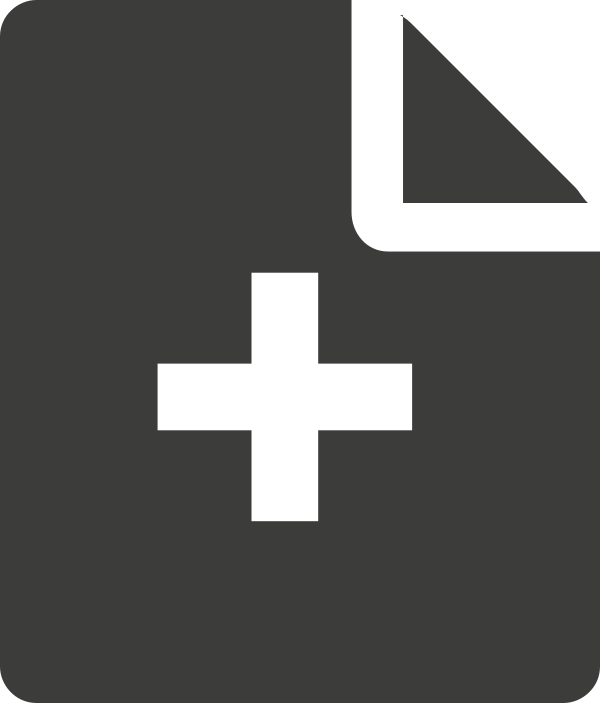
***In silico* screening-based discovery of inhibitors against glycosylation proteins dysregulated in cancer**

**Michael Russelle S. Alvarez, Sheryl Joyce B. Grijaldo, Ruel C. Nacario, Jomar F. Rabajante, Francisco M. Heralde III, Carlito B. Lebrilla & Gladys C. Completo**

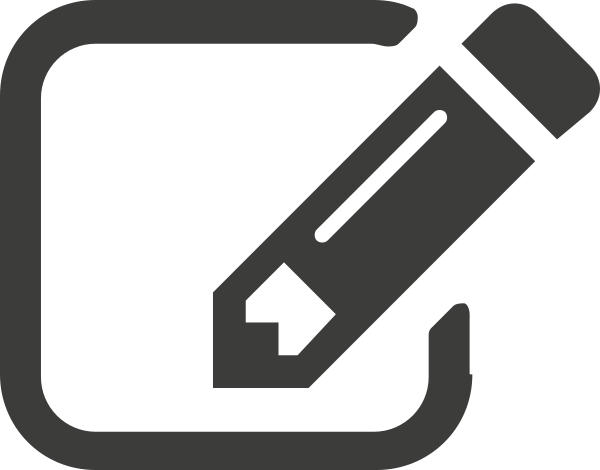
**To cite this article:** Michael Russelle S. Alvarez, Sheryl Joyce B. Grijaldo, Ruel C. Nacario, Jomar

F. Rabajante, Francisco M. Heralde III, Carlito B. Lebrilla & Gladys C. Completo (2022): *In silico* screening-based discovery of inhibitors against glycosylation proteins dysregulated in cancer, Journal of Biomolecular Structure and Dynamics, DOI: [10.1080/07391102.2021.2022534](https://www.tandfonline.com/action/showCitFormats?doi=10.1080/07391102.2021.2022534)

**To link to this article:** <https://doi.org/10.1080/07391102.2021.2022534>

[View supplementary material](https://www.tandfonline.com/doi/suppl/10.1080/07391102.2021.2022534) 

 Published online: 06 Jan 2022.

 [Submit your article to this journal](https://www.tandfonline.com/action/authorSubmission?journalCode=tbsd20&show=instructions)

[View related articles](https://www.tandfonline.com/doi/mlt/10.1080/07391102.2021.2022534) 

[View Crossmark data](http://crossmark.crossref.org/dialog/?doi=10.1080/07391102.2021.2022534&domain=pdf&date_stamp=2022-01-06)

Full Terms & Conditions of access and use can be found at <https://www.tandfonline.com/action/journalInformation?journalCode=tbsd20>

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS



<https://doi.org/10.1080/07391102.2021.2022534>



In silico screening-based discovery of inhibitors against glycosylation proteins dysregulated in cancer

# Michael Russelle S. Alvareza,b, Sheryl Joyce B. Grijaldoa, Ruel C. Nacarioa, Jomar F. Rabajantec, Francisco M.

Heralde IIId, Carlito B. Lebrillae and Gladys C. Completoa

aInstitute of Chemistry, University of the Philippines Los Ban~os, Los Ban~os, Laguna, Philippines; bCollege of Arts and Sciences, Isabela State University, Echague, Isabela, Philippines; cInstitute of Mathematical Sciences and Physics, University of the Philippines Los Ban~os, Los Ba~nos, Laguna, Philippines; dMolecular Diagnostics and Cellular Therapeutics Laboratory, Lung Center of the Philippines, Quezon City, Philippines; eDepartment of Chemistry, University of California Davis, Davis, California, USA

Communicated by Ramaswamy H. Sarma

ARTICLE HISTORY

ABSTRACT

Targeting enzymes associated with the biosynthesis of aberrant glycans is an under-utilized strategy in discovering potential inhibitors or drugs against cancer. The formation of cancer-associated glycans is mainly due to the dysregulated expression of glycosyltransferases and glycosidases, which play crucial roles in maintaining cellular structure and function. We screened a database of more than 14,000 com- pounds consisting of natural products and drugs for inhibition against four glycosylation enzymes - Alpha1-6FucT, ST6Gal1, ERMan1, and GlcNAcT-V. The top inhibitors identified against each enzyme were subsequently analyzed for potential binding against all four enzymes. *In silico* screening results show several promising candidates that could potentially inhibit all four enzymes: (1) Amb20622156

GlcNAcT-V: -7.2 kcal/mol], (2) Amb22173588 (1,2-dihydrotanshinone I) [ERMan1: -9.3 kcal/mol; Alpha1-

(demethylwedelolactone) [ERMan1: -9.3 kcal/mol; Alpha1-6FucT: -7.3 kcal/mol; ST6Gal1: -8.4 kcal/mol;

nol B) [ERMan1: -9.3 kcal/mol; Alpha1-6FucT: -6.0 kcal/mol; ST6Gal1: -9.8 kcal/mol; GlcNAcT-V:

6FucT: -6.1 kcal/mol; ST6Gal1: -9.2 kcal/mol; GlcNAcT-V: -7.9 kcal/mol], and (3) Amb22173591 (tanshi-

form non-covalent bonding interactions with key active site residues in each enzyme, suggesting crit- ical target residues in the four enzymes’ active sites. Furthermore, pharmacokinetic property prediction analysis using pkCSM indicates that all of these inhibitors have good ADMETox properties (i.e., log P < 5, Caco-2 permeability > 0.90, intestinal absorption > 30%, skin permeability>-2.5, CNS permeabil-

ity <-3, maximum tolerated dose < 0.477, minnow toxicity<-0.3). The *in silico* docking approach to

glycosylation enzyme inhibitor prediction could help guide and streamline the discovery of novel

inhibitors against enzymes involved in aberrant protein glycosylation.

-7.7 kcal/mol]. Drug-enzyme active site residue interaction analyses show that the putative inhibitors

Received 25 August 2021

Accepted 20 December 2021

KEYWORDS

Glycosylation; *in silico* docking; glycosidases; glycosyltransferases; ADMETox

# Introduction

The typical timeline for drug discovery and development process takes between 10-15 years and costs $314 million to

$2.8 billion (Mohs & Greig, [2017](#_bookmark99); Wouters et al., [2020](#_bookmark117)). This process is generally viewed as time-consuming, labor-inten- sive, expensive, and has low success rates (Green, [2003](#_bookmark61); Pozzan, [2006](#_bookmark91)). Therefore, computer-aided drug discovery and development helps in reducing the time and costs by lever- aging computational power, knowledge of the chemical and biological information of the ligands and drug targets, and design of in silico filters such as ADMETox) (Kapetanovic, [2008](#_bookmark70)). Large compound libraries, such as those containing drugs like ChEMBL (Mendez et al., [2019](#_bookmark97)), PubChem (H€ahnke et al., [2018](#_bookmark62)), and ChemSpider (Pence & Williams, [2010](#_bookmark111)), serve as excellent starting points since these databases contain a wide range of small molecule compounds (Sorokina & Steinbeck, [2020](#_bookmark103)). ChEMBL, a product of the European

Bioinformatics Institute, is particularly attractive due to its focus on experimentally-elucidated drugs and drug-like com- pounds. Industrial catalogs are also a good source of chem- ical libraries due to ease of access for further *in vitro* validation after *in silico* screening and hit identification. One such example is the natural product catalog from Ambinter- Greenpharma (<http://www.ambinter.com/>). Another approach is to utilize known drug-protein interactions from databases such as DrugBank (<https://go.drugbank.com/>), Comparative Toxicogenomics Database (<http://ctdbase.org/>), STITCH (<http://stitch.embl.de/>), GeneCards ([https://www.genecards.](https://www.genecards.org/) [org/](https://www.genecards.org/)), Drug Gene Interaction database ([http://www.dgidb.](http://www.dgidb.org/) [org/](http://www.dgidb.org/)), and Protein Databank (<https://www.rcsb.org/>). Identifying drug-protein interactions helps provide informa- tion on a drug’s pharmacology, efficacy as well as strategies to identify drug candidates against a target protein (Choudhary & Singh, [2019](#_bookmark67)).

CONTACT Gladys Cherisse J. Completo  [gjcompleto@up.edu.ph](mailto:gjcompleto@up.edu.ph) Institute of Chemistry, University of the Philippines Los Ban~os, Los Ba~nos, Laguna, Philippines

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07391102.2021.2022534>.

© 2022 Informa UK Limited, trading as Taylor & Francis Group

2 M. R. S. ALVAREZ ET AL.

Tumor cells display a wide range of glycosylation altera- tions than normal cells (Pinho & Reis, [2015](#_bookmark84)). Glycosylation, the addition of glycans to proteins and lipids, occurs in the endoplasmic reticulum (ER) or Golgi apparatus and is cata- lyzed by glycosyltransferases and glycosidases. These enzymes form glycans in a series of steps dependent on pro- tein substrate bioavailability, enzyme activity, altered enzyme location within subcellular components, and levels of gene transcription (Reily et al., [2019](#_bookmark98)). With the differential expres- sion of glycosylation enzymes between normal cells and can- cer cells, targeting the enzymes involved in the biosynthesis of aberrant glycans can be used as a strategy to discover anticancer drugs. Two principal mechanisms contributing to the formation of these tumor-associated glycans are the incomplete synthesis and neo-synthesis processes (Kannagi et al., [2008](#_bookmark68)). During the early stages of cancer, the incom- plete synthesis process is a consequence of impairing the normal synthesis of complex glycans expressed in epithelial cells leading to biosynthesis of truncated structures such as sialyl Tn (STn) (Munkley, [2016](#_bookmark104)). Moreover, neo-synthesis is more commonly observed in advanced cancer stages that involve the induction of cancer-associated genes involved in the expression of glycan cancer biomarkers like sialyl Lewisa (SLea) and SLex (Chen et al., [2016](#_bookmark66); Cohen et al., [2019](#_bookmark71)). In add- ition, several cancer-associated glycosylation changes such as b1,6 branching, sialyl Lewis antigens, a2,6-sialylated lactos- amine, T, Tn, and sialyl-Tn antigens, and gangliosides/glyco- sphingolipids have also been documented (Dall’Olio et al., [2012](#_bookmark79)).

The role of glycosyltransferases in aberrant glycosylation of cancer is well documented, and some of these enzymes were considered cancer biomarkers (Meany & Chan, [2011](#_bookmark95)). A glyco- syltransferase used as a biomarker is UDP-N-acetyl-D-glucosa- mine-N-acetylglucosamine transferase V (GlcNAcT-V, Gene name: MGAT5), which catalyzes b1-6 branching of N-glycans. Increased b1-6 branching, due to GlcNAcT-V overexpression, has been observed in breast cancer (Handerson et al., [2005](#_bookmark65)). Sialyltransferases are another example of glycosyltransferases that are abnormally expressed in cancers and are implicated in carcinogenesis, progression, and metastasis (Burchell et al., [1999](#_bookmark63); Picco et al., [2010](#_bookmark81); Recchi et al., [1998](#_bookmark96)). The a2,6-sialylated lactosamine (Sia6LacNAc) is the product of b-galactoside a2,6- sialyltransferase (ST6Gal1, Gene Name: SIAT1) (Weinstein et al., [1987](#_bookmark117)), and its expression is altered in several cancers such as colon, stomach, and ovarian cancer (F Dall’Olio & Chiricolo, [2001](#_bookmark78)). Core fucosylation is also observed in several cancers (Pinho & Reis, [2015](#_bookmark84)). It involves adding a1,6-fucose to the innermost GlcNAc residue of N-glycans through Fuc-TVIII (Alpha1-6FucT, Gene Name: FUT8). Overexpression is observed in several cancers, including lung cancer (Liu et al., [2011](#_bookmark82)). In colorectal cancer, the aberrant glycosylation stems from the defects in the glycosyltransferase genes with identified inacti- vating germline and somatic mutations in the gene encoding for GALNT12 (Polypeptide N-acetylgalactosaminyltransferase 12), B3GNT2 (b-1,3-N Acetylglucosaminyltransferase 2), B4GALT2 (b-1,4-Galactosyltransferase 2), and ST6GALNAC2 (a-N-Acetylgalactosaminide a-2,6-Sialyltransferase 2), predom- inantly mapped to these enzymes’ respective catalytic

domains (Venkitachalam & Guda, [2017](#_bookmark114)). In breast cancer, the increased core fucosylation of epidermal growth factor recptor (EGFR) was associated with increased dimerization and phos- phorylation, resulting in increased EGFR-mediated signaling and subsequent promotion of tumor growth (Potapenko et al., [2010](#_bookmark87)). In lung cancer patients, aberrant glycosylation is correlated to aberrant expression of glycosylation enzymes as well. Gene-expression analysis of lung tissue sections from smokers and non-smokers showed significantly upregulated MAN1A2, MAN2A1, MGAT2, MGAT4B, B4GALT2, FUT2, FUT3,

FUT6, and FUT8 while several enzymes, MAN1A1, MAN1C1, MAN2A2, MGAT1, MGAT3, and FUT1, were significantly down- regulated (Landi et al., [2008](#_bookmark76); Ruhaak et al., [2018](#_bookmark101)).

Thus, in the current study, we prepared a ligand database from several open-access chemical databases and subsequently screened the resulting 14,777 compounds against the crystal structures of Alpha1-6FucT, ST6Gal1, ERMan1, and GlcNAcT-V. These glycosylation enzymes were selected based on their reported role in the aberrant glycosylation in cancer and their contribution to disease progression (Burchell et al., [1999](#_bookmark63); F Dall’Olio & Chiricolo, [2001](#_bookmark78); Handerson et al., [2005](#_bookmark65); Landi et al., [2008](#_bookmark76); Liu et al., [2011](#_bookmark82); Picco et al., [2010](#_bookmark81); Pinho & Reis, [2015](#_bookmark84); Potapenko et al., [2010](#_bookmark87); Recchi et al., [1998](#_bookmark96); Ruhaak et al., [2015](#_bookmark100); Weinstein et al., [1987](#_bookmark117)). The top inhibitors against each enzyme were selected, and the protein-ligand interactions between the inhibitors and protein active sites were assessed to identify key active site amino acid residues responsible for ligand binding. In addition, we analyzed the ability of the compounds to poten- tially inhibit multiple glycosylation enzymes, providing a multi- targeting inhibition platform. Finally, we filtered these top inhib- itors for druggability based on favorable ADMETox properties, as predicted by pkCSM (Pires et al., [2015](#_bookmark88)).

# Materials and methods

## *Database and ligand preparation*

We prepared a ligand database by downloading the struc- ture data files (.sdf) from several online databases: 2,749 Phase 4 drugs from ChEMBL database ([https://www.ebi.ac.](https://www.ebi.ac.uk/chembl/) [uk/chembl/](https://www.ebi.ac.uk/chembl/)), 11,622 natural product compounds from the Ambinter catalog (<http://www.ambinter.com/#catalog>), and 185 compounds predicted to bind or interact with glycosyla- tion enzymes according to DrugBank ([https://go.drugbank.](https://go.drugbank.com/) [com/](https://go.drugbank.com/)), Comparative Toxicogenomics database ([http://](http://ctdbase.org/) [ctdbase.org/](http://ctdbase.org/)), STITCH database (<http://stitch.embl.de/>), GeneCards (<https://www.genecards.org/>), Drug Gene Interaction database (<http://www.dgidb.org/>), and Protein Databank (<https://www.rcsb.org/>). Additionally, an in-house database composed of 221 natural product compounds pre- viously isolated from *Lansium parasiticum* (Manosroi et al., [2012](#_bookmark89); Marfori et al., [2015](#_bookmark90)), *Mangifera indica* (D[Zbreve]Ami'c et al., [2010](#_bookmark74); Masibo & He, [2008](#_bookmark93); Nu'n~ez Sell'es et al., [2002](#_bookmark107); Pino & Mesa, [2006](#_bookmark85)), and *Annona muricata* (Coria-T'ellez et al., [2018](#_bookmark72)) were prepared by drawing using MarvinSketch and subsequently converted to 3D structures. All 14,777 com- pounds were then loaded onto PyRx (Dallakyan & Olson, [2015](#_bookmark77)) and minimized using the Universal Force Field (Rappe et al., [1992](#_bookmark94)) as implemented in Open Babel (O’Boyle et al.,

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS  3

[2011](#_bookmark109)). These compounds were subsequently screened for this study ([Supplementary Table 1](#_bookmark30)).

Leu372

Lys376

Hbond

Hbond

Hbond

Hbond

Pi-alkyl

Hbond

## *Protein preparation*

The enzymes GlcNAcT-V, ERMan1, ST6Gal1, and Alpha1-6FucT were selected as drug targets for this study due to their reported roles in cancer progression. The enzyme GlcNAcT-V (PDB ID: 5ZIC, 2.10 Ð, Nagae et al., [2018](#_bookmark105)) was downloaded as a complex with its acceptor sugar, 2-acetamido-2-deoxy-beta-D- glucopyranose-(1-2)-6-thio-alpha-D-mannopyranose-(1-6)-

Tyr369

His370

Hbond

Pi-Pi

T shaped

beta-D-mannopyranose. ERMan1 (PDB ID: 1X9D, 1.41 Ð, Karaveg et al., [2005](#_bookmark73)) was also downloaded as a complex with a thio-disaccharide substrate analog, alpha-D-mannopyranose- (1-2)-methyl 2-thio-alpha-D-mannopyranoside. ST6Gal1 (PDB ID: 4JS2, 2.30 Ð, Kuhn et al., [2013](#_bookmark75)) was also downloaded as a complex with cytidine monophosphate. While the structure of the human Alpha1-6FucT (PDB ID: 2de0, Ihara et al., [2007](#_bookmark69)) was designed using homology modeling from *Caenorhabditis ele- gans* POFUT1 (PDB ID: 3ZY6, 1.91 Ð, Lira-Navarrete et al., [2011](#_bookmark83)) in complex with GDP-fucose using SWISS-MODELLER (Waterhouse et al., [2018](#_bookmark115)) and SchrodingerTM with root mean square deviation (RMSD) value of 1.144. The protein structures were subsequently prepared for docking using the Dockprep protocol in Chimera (Pettersen et al.,[2004](#_bookmark112)), which involved energy minimization in 200 steps using the Amberff14SB force field. The prepared protein structures were then loaded in PyRx as macromolecule receptors.

Table 1. Interactions of key enzyme amino acid residues and the top 11 compounds screened against ST6Gal1.

Binding energy

Gly324

Glu342

Cys353

Hbond

Tyr354

Hbond

[Cys364](#_bookmark69)

Thr365

Hbond

Ala368

Hbond

Hbond

Hbond

Amide-Pi stacked

Hbond

Hbond

Pi-sulfur

Amide-Pi stacked

Hbond Hbond

Alkyl

Hbond

Halogen

(fluorine)

Halogen (fluorine)

Halogen (fluorine)

Alkyl

## *In silico screening*

Ser323

Hbond Hbond

Hbond Hbond

Hbond Hbond

The *in silico* screening methods were performed in PyRx (Dallakyan & Olson, [2015](#_bookmark77)) using the AutoDock Vina docking protocol (Trott & Olson, [2010](#_bookmark110)) at exhaustiveness (E, parameter for comprehensive search in AutoDock Vina) level 8. Before docking, the ligands complexed with the respective enzymes in the PDB crystal structure (or GDP-fucose, in the case of human Alpha1-6FucT) were docked to validate the docking protocol, specifically the grid box parameters. The docking protocol was validated when the docked ligand and crystal ligand had an all-atom RMSD, a measure of the quality of reproduction, less than 1.5 ([Supplementary Table 2](#_bookmark39)). After docking validation, all the 14,777 database compounds were screened against each of the four enzymes. The binding ener- gies and all conformations produced were exported as .pdbqt files for further analysis. The compounds were ranked accord- ing to the Vina-predicted binding energy (kcal/mol). The top binding molecules against each enzyme were visualized for residue interactions with the target enzyme using Discovery StudioTM. Compound cross-reactivity (i.e., binding to multiple enzyme targets) was analyzed by evaluating each compound’s binding energy against the individual enzyme and the data generated was represented as a heatmap.

Ligand

CMP (cytidine monophosphate)

Amb23603897

(Limonexin)

Amb22584490

(Evodol) Amb10549471 Amb10549969

(Ginkgolide A) Amb17621731 Amb19746905

(kcal/mol)

-8

-11.1

-10.8

Ser188

Ser189

Hbond

Ala190

Asn212

Hbond

[Asn233](#_bookmark109)

Ser322

Hbond

Amide-Pi

stacked Hbond

Pi-sigma

Hbond

[Hbond](#_bookmark109)

-10.6

-10.6

Hbond

Hbond

-10.5

Hbond Hbond

Hbond Hbond

-10.5

-10.5

Pi-alkyl

Pi-sigma

Amb22584663

(Rutaevin) Amb28533068 Amb29844395

(Grantianine) Amb29844418 CKEMBL3989866

(Bictegravir)

Amide-Pi

stacked Hbond Hbond

Pi-sigma

Hbond

[Hbond](#_bookmark109)

-10.5

Hbond

-10.4

Hbond

-10.4

Hbond Pi-alkyl

Hbond

-10.4

Pi-sigma

Hbond

## *ADMETox analysis*

The ADMETox analysis and druggability predictions of the top 10 compounds against each enzyme target were made

4 M. R. S. ALVAREZ ET AL.

Table 2. Interaction of key enzyme amino acid residues and the top 10 compounds screened against GlcNAcT-V.

Binding energy

Ligand

2-acetamido-2-deoxy-beta- D-glucopyranose-(1-2)-

6-thio-alpha-D-mannopyranose- (1-6)-beta-D-mannopyranose

Amb31207291

Benzoperylene

(kcal/mol)

-8.0

Leu276

Phe283

Trp301

Asp352

Ile353

Leu377

Asp378

Hbond

Ser379

Hbond

Phe380

Pro400

Trp401

Hbond

Phe550

VDW

Lys554

Hbond

Pro555

-10.2

Hbond Pi-alkyl

VDW

Hbond Pi-pi Stacked

-9.8

-9.6

Pi-alkyl

Pi-pi

stacked

Amb23604039

Hbond

Hbond

Amb23604160 CHEMBL2103870

-9.6

Pi-alkyl

Pi-sigma

Pi-alkyl

Hbond Hbond

-9.6

Pi-alkyl

Halogen

(fluorine)

Pi-alkyl

VDW

Hbond

Hbond

Hbond Hbond

Hbond

using the pkCSM server (Pires et al., [2015](#_bookmark88)). The compounds were filtered based on the following druggability criteria: Lipinski (Lipinski et al., [2001](#_bookmark80)), Ghose (Ghose et al., [1999](#_bookmark60)), Veber (Veber et al., [2002](#_bookmark113)), Egan (Egan et al., [2000](#_bookmark56)), Muegge (Muegge et al., [2001](#_bookmark102)), and Bioavailability score (Martin, [2005](#_bookmark92)).

# Results and discussion

## *Database and ligand preparation*

A schematic diagram of the *in silico* screening and docking strategy is shown in [Supplementary Figure 1](#_bookmark45). First, a com- pound library was constructed consisting of the 14,777 com- pounds from Ambinter, ChEMBL, DRUGBANK, CTD, STITCH, DGIdb, GeneCards, and PDB. The 11,622 compounds from the Ambinter catalog were natural product compounds obtained from Ambinter GreenPharma, one of the freely available online natural product structure databases (Sorokina & Steinbeck, [2020](#_bookmark103)). From the ChEMBL database, 2,749 compounds were selected by filtering only compounds with known Phase 4 clinical data (Mendez et al., [2019](#_bookmark97)). Next, compounds that are known to interact with proteins involved in glycosylation (glycosylation interactors) were manually searched from several gene-drug interaction data- bases having reported interactions with the following glyco- sylation genes: Alpha1-6FucT, FUT7, FUT3, FUT4, ST6Gal1, MAN1A1, MAN1A2, MGAT1, MGAT2, GlcNAcT-V, B4GALT1,

Hbond Pi-pi

Stacked Hbond

Pi-pi Stacked

VDW

Hbond Hbond

Pi-alkyl

Hbond

Hbond

and B4GALT2 (Davis et al., [2021](#_bookmark57); Freshour et al., [2021](#_bookmark58); Stelzer et al., [2016](#_bookmark106); Szklarczyk et al., [2019](#_bookmark108); Wishart et al., [2018](#_bookmark116)). These glycosylation genes were selected due to their being upregulated in specific cancer types (Landi et al., [2008](#_bookmark76); Ruhaak et al., [2018](#_bookmark101)). Moreover, the N-glycans synthesized by these glycosylation proteins are also associated with cancer progression (Ruhaak et al., [2018](#_bookmark101)).

VDW

Pi-alkyl

Hbond

[Hbond](#_bookmark101)

The glycosylation enzymes ST6Gal1, ERMan1, GlcNAcT-V, and Alpha1-6FucT were selected based on their reported role in aberrant glycosylation in cancer contributing to dis- ease progression (Burchell et al., [1999](#_bookmark63); Dall’Olio & Chiricolo, [2001](#_bookmark78); Handerson et al., [2005](#_bookmark65); Landi et al., [2008](#_bookmark76); Liu et al., [2011](#_bookmark82); Picco et al., [2010](#_bookmark81); Pinho & Reis, [2015](#_bookmark84); Potapenko et al., [2010](#_bookmark87); Recchi et al., [1998](#_bookmark96); Ruhaak et al., [2015](#_bookmark100); Weinstein et al., [1987](#_bookmark117)). Although the Alpha1-6FucT protein was downloaded in its apo-form (PDB ID: 2de0), a homology model of its holo-form was constructed using *C. elegans* Alpha1-6FucT (PDB ID: 3zy) as a protein template. The 14,777 compounds were docked individually against each target enzyme after performing docking validations ([Supplementary Table 2](#_bookmark39)). Most of the compounds bound well against ST6Gal1 and GlcNAcT-V, with the best compound determined to have

-9.5

-9.5

-9.5

Pi-sigma

-9.3

Pi-alkyl

-9.3

binding affinities of -11.1 kcal/mol and -9.8 kcal/mol against

ST6Gal1 and GlcNAcT-V, respectively ([Supplementary](#_bookmark46) [Figure 2](https://doi.org/10.1080/07391102.2021.2022534)).

## *ST6Gal1 screening*

Amb22584370 Amb22801070

Amb15769953 Amb23603914 Amb18511396

For the virtual screening of ST6Gal1 (PDB ID: 4JS2), the pro- tein was prepared using standard protein preparation proto- cols (i.e. removal of solvents, energy minimization). The

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS  5

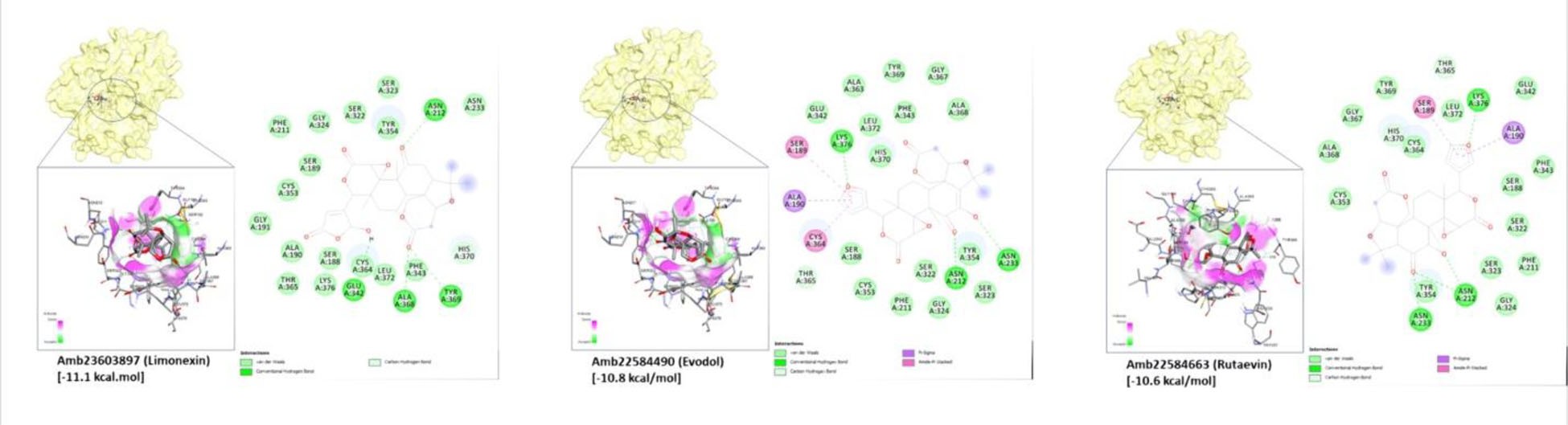


Figure 1. Docking conformation and amino acid residue interactions of the top 3 compounds, (Amb23603897 (limonexin), Amb22584490 (evodol), and Amb22584663 (rutaevin), screened against ST6Gal1. The drug-residue interactions are color-coded based on the type of interactions.

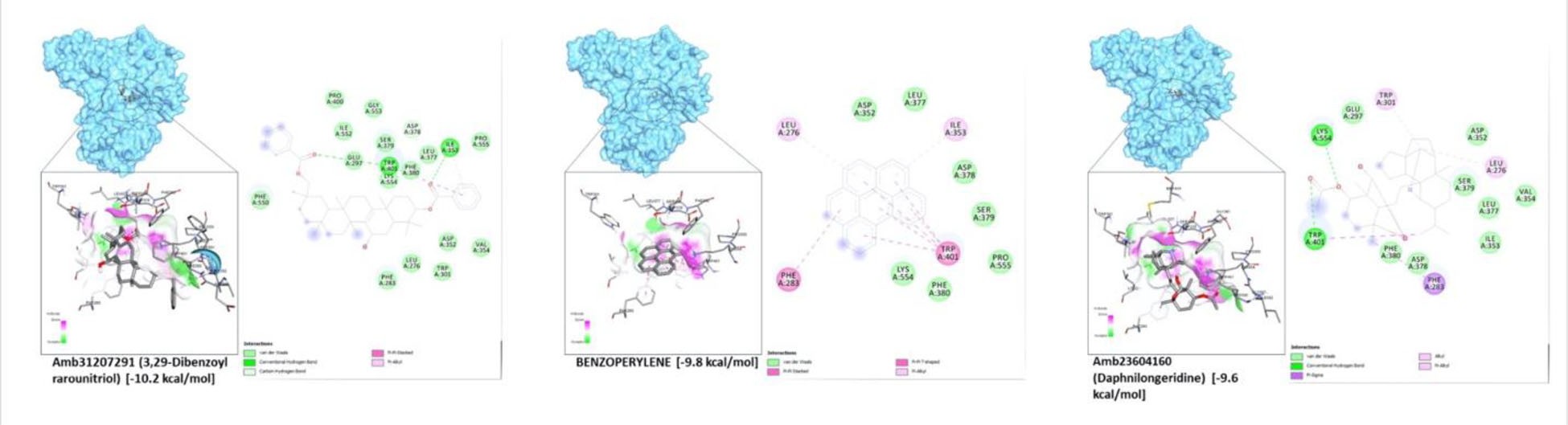


Figure 2. Docking conformation and amino acid residue interactions of the top 3 compounds screened against GlcNAcT-V: Amb31207291 (3,29-dibenzoylrarouni- triol), benzoperylene, and Amb23604160 (Daphnilongeridine). The drug-residue interactions are color-coded based on the type of interactions.

available 3D crystal structure of ST6Gal1 was co-crystallized with cytidine monophosphate (CMP). Thus, CMP was used as the positive control for docking validation using the grid box with the following coordinates (4.6524, 52.3276, 37.3177) and dimensions (9.1532 Ð, 8.2193 Ð, 14.2562 Ð). The docking val- idation showed that the docked CMP molecule had very similar conformation (RMSD 0.231) and interactions with the protein. The compound library was screened against ST6Gal1, with a binding energy frequency shown below. Several compounds from the library gave significantly higher binding affinity than the positive control CMP (binding affin- ity 8.0 kcal/mol). Specifically, the following are the top compounds that showed higher binding affinity to ST6Gal1 compared to CMP after virtual screening: Amb23603897 (limonexin) [-11.1 kcal/mol], Amb22584490 (evodol) [-10.8 kcal/mol], Amb10549471 [-10.6 kcal/mol], Amb22584663 (Rutaevin) [-10.6 kcal/mol], Amb17621731 [-10.5 kcal/mol], Amb19746905 [-10.5 kcal/mol], Amb28533068 [-10.5 kcal/mol], CHEMBL39898 (Bictegravir) [-10.5 kcal/mol], Amb10549969 (Ginkgolide A) [-10.4 kcal/mol], Amb29844395 (Grantianine) [-10.4 kcal/mol], and Amb299844418 [-10.4 kcal/mol] ([Figure 1](#_bookmark45); [Supplementary](#_bookmark47) [Figure 3](https://doi.org/10.1080/07391102.2021.2022534)).

¼

¼ -

Comparison of the docking poses and amino acid interac-

tions of the top ligands with the positive control, CMP, allowed the identification of several amino acid residues in the active site as important sites for strong binding by poten- tial inhibitors. Several top ligands, including CMP, have H- bonding interactions with Ser189, Asn212, and Lys376. Additionally, Amb23603897 (limonexin) [-11.1 kcal/mol] had additional H-bond interactions with Glu342, Ala368, and

Tyr369 ([Table 1](#_bookmark30)). Whereas CMP formed H-bonding interaction with Cys353 (Kuhn et al., [2013](#_bookmark75)), Amb19746905 [-10.5 kcal/ mol], and CHEMBL39898 (Bictegravir) [-10.5 kcal/mol] formed pi-sulfur and halogen (fluorine) interactions with the Cys353 residue as well. Kuhn and his coworkers compared the bind- ing of CMP in ST6Gal1 with that of ST3/8 from *Campylobacter jejuni* and ST3Gal-1 from *Sus scrofa* and their results showed that the A-phosphate of CMP has H-bonding interaction with a tyrosine residue in the active site (Kuhn et al., [2013](#_bookmark75)). In ST6Gal1, the tyrosine residue is identified as Tyr354. Although all the top ligands did not interact directly with Tyr 354, all of

the ligands are in the vicinity of Tyr354. Moreover, the 30-

hydroxyl group of CMP forms a hydrogen bonding interaction with Gly324 in ST6Gal1. And some of the screened com- pounds, Amb10549471 [-10.6 kcal/mol], Amb28533068 [-10.5 kcal/mol], Amb29844395 (Grantianine) [-10.4 kcal/mol], and CHEMBL39898 (Bictegravir) [-10.5 kcal/mol], similarly formed H-bonding interaction with the Gly234 residue. A plausible reaction mechanism proposed by Kuhn and cow- orkers shows that His370 could act as the catalytic base for

the deprotonation of the 60-hydroxyl group of the acceptor N-

glycan, leading to SN2 attack of the C2 atom of Neu5Ac (Kuhn et al., [2013](#_bookmark75)). Additionally, CHEMBL39898 (Bictegravir) [-10.5 kcal/mol] was also shown to form Pi-Pi T-shaped interac- tions with the His370 residue from our binding studies.

## *GlcNAcT-V screening*

For the virtual screening of GlcNAcT-V (PDB ID: 5ZIC), the protein was prepared using standard protein preparation

6  M. R. S. ALVAREZ ET AL.

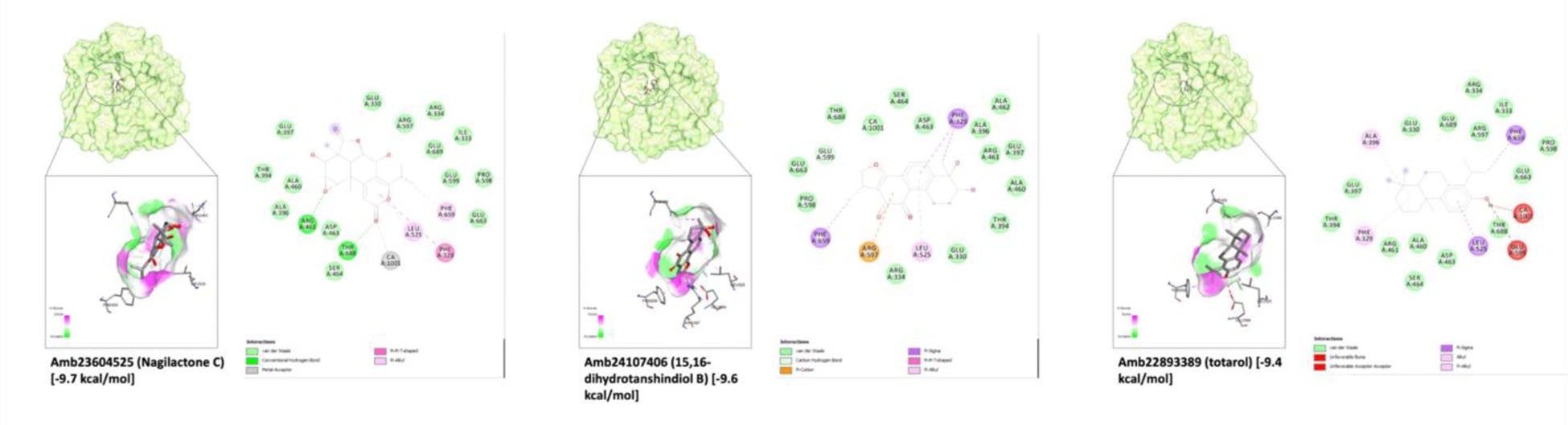


Figure 3. Docking conformation and amino acid residue interactions of the top 3 compounds screened against ERMan1: Amb23604525 (Nagilactone C), Amb24107406 (15,16-dihydrotanshindiol B), Amb22893389 (totarol). The drug-residue interactions are color-coded based on the type of interactions.

protocols (removal of solvents, energy minimization). The available 3D crystal structure of GlcNAcT-V was co-crystal- lized with a sulfated-acceptor sugar (Man2GlcNAc1). This acceptor sugar was used as the positive control in the dock- ing validation using the gridbox with the following coordi-

nates (29.7431, -15.1393, 31.3969) and dimensions (8.6388

Ð, 12.1308 Ð, 9.6086 Ð). Using these docking parameters, the compound library was screened against GlcNAcT-V. Compared to the positive control (binding affinity ¼

-6.9 kcal/mol), several compounds from the library gave sig-

nificantly higher binding affinities. Specifically, these com- pounds are the Amb31207291 (3,29-dibenzoyl karounitriol) [-10.2 kcal/mol], Benzoperylene [-9.8 kcal/mol], Amb23604039 (3,21-dihydroxy-14-serraten-16-one) [-9.6 kcal/mol], Amb23604160 (daphnilongeridine) [-9.6 kcal/mol], CHEMBL2103870 (lumacaftor) [-9.6 kcal/mol], Amb22584370 (Absinthiin) [-9.5 kcal/mol], Amb22801070 (Strychnine) [-9.5 kcal/mol], Amb15769953 [-9.3 kcal/mol], Amb23603914 (Neoprzewaquinone A) [-9.3 kcal/mol], and Amb18511396 [-9.3 kcal/mol] ([Figure 2](#_bookmark46); [Supplementary Figure 4](#_bookmark48)).

Further analysis of the docking pose of the top binding

compounds showed interesting interactions with key active site residues ([Table 2](#_bookmark39)). Most of the top ligands, including the positive control, formed hydrogen bonding interaction with Trp401. While benzoperylene [-9.8 kcal/mol], CHEMBL2103870 (lumacaftor) [-9.6 kcal/mol], and Amb22801070 (Strychnine) [-9.5 kcal/mol] formed pi-pi stacking interactions with Trp401 instead of hydrogen bonds. This is presumably due to the presence of conjugated double bonds in the ligands which are situated near the Trp401 residue. CHEMBL2103870 (lumacaf- tor) [-9.6 kcal/mol] also formed Pi-alkyl interactions with Phe380. Nagae and coworkers suggested that these two aro- matic residues, Phe380 and Trp401, are interacting with the acceptor sugar, that involved Trp401 restraining the conform- ation of the A1,6-branch (Nagae et al., [2018](#_bookmark105)). Most of the top ligands also formed hydrogen bonding interaction with Lys554. In the crystal structure, this residue interacts with the acceptor sugar. The three amino acid residues - Phe380, Trp401, and Lys554 - are also found in the acceptor substrate binding site for *GlcNAcT-V*, suggesting that these residues are relevant in acceptor sugar recognition (Nagae et al., [2018](#_bookmark105)). Thus, the top ligands that were selected in this study are the ones that form strong binding interactions with these amino acid residues.

## *ERMan1 screening*

For the virtual screening of ERMan1 (PDB ID: 1X9D), the pro- tein was prepared following standard protein preparation pro- tocols (removal of solvents, energy minimization). The 3D crystal structure available was co-crystallized with its acceptor substrate (Man9GlcNAc2). The acceptor substrate was also used as the positive control for docking validation using the gridbox with the following coordinates (2.5758, 6.3397, 0.6598) and dimensions (6.4456 Ð, 10.5668 Ð, 6.2261 Ð). The docking validation showed that the redocked molecule had very similar conformation (RMSD 0.587) and interactions with the protein. Compared to the positive control (binding affinity 7.6 kcal/mol), several compounds from the library gave significantly higher binding affinities ([Figure 3](#_bookmark47); [Supplementary Figure 5](https://doi.org/10.1080/07391102.2021.2022534)). Specifically, these compounds are the Amb23438478 [-9.7 kcal/mol], Amb23604525 (Nagilactone C) [-9.7 kcal/mol], Amb24107406 (15,16-dihydrotanshindiol B) [-9.6 kcal/mol], Amb22893389 (totarol) [-9.4 kcal/mol], Amb23604091 (totaradiol) [-9.4 kcal/mol], Amb20622156 (demethylwedelolactone) [-9.3 kcal/mol], Amb22173588 (1,2- dihydrotanshinone I) [-9.3 kcal/mol], Amb22173591 (tanshinol B) [-9.3 kcal/mol], Amb22173639 (1,2-didehydrocryptotanshi- none) [-9.3 kcal/mol], Amb23603874 (scholaricine) [-9.3 kcal/ mol], and Amb24107366 (epinodosin) [-9.3 kcal/mol].

-

¼

¼ -

Comparison of the binding interactions of the top ligands with ERMan1 allowed the identification of important amino acid residues that commonly bind to these ligands ([Figure](#_bookmark47) [3](#_bookmark47)). Most of the top ligands, including the positive control,

showed hydrogen bonding interaction with Thr688. This resi- due was also found to directly coordinate with Ca2þ ion, considered as a cofactor of mannosidase, in the crystal struc- ture (Karaveg et al., [2005](#_bookmark73)). Karaveg and coworkers also iden- tified several potential catalytic base residues; Asp463, Glu330, and Glu599 with Arg334, Glu330, His524, and Glu599

potentially aiding in the acid-base hydrolysis of the substrate (Karaveg et al., [2005](#_bookmark73)). During our docking experiments, Arg334 formed hydrogen bonding interactions with the ligands Amb23603874 (scholaricine) [-9.3 kcal/mol] and Amb24107366 (epinodosin) [-9.3 kcal/mol]. The residue Asp463 mainly showed Pi-anion interactions with Amb23604525 (Nagilactone C) [-9.7 kcal/mol] and Amb22173588 (1,2-dihydrotanshinone I) [-9.3 kcal/mol], while Glu330 formed Pi-anion interactions with Amb23604525 (Nagilactone C) [-9.7 kcal/mol], Amb22173588 (1,2-

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS  7

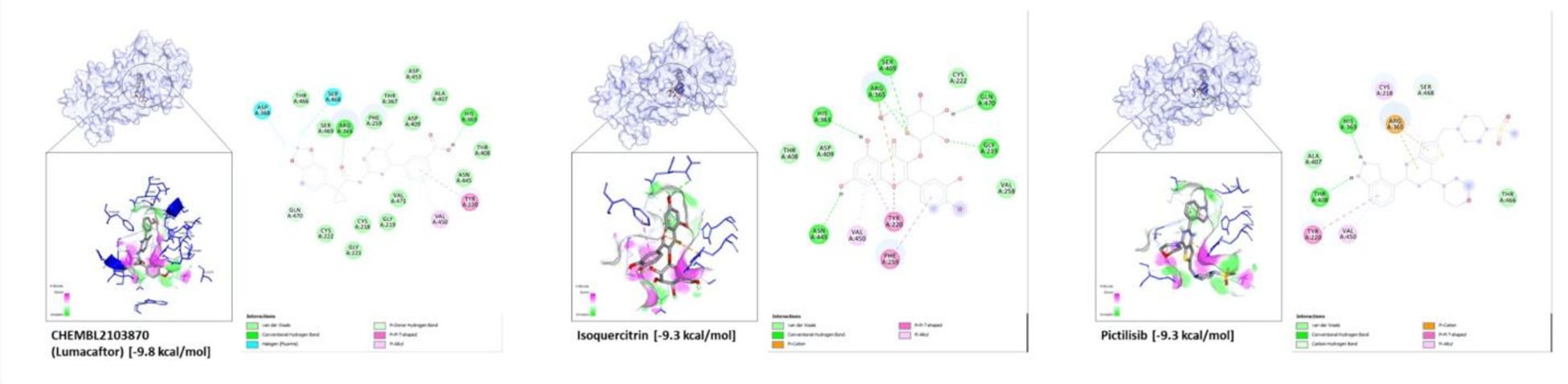


Figure 4. Docking conformation and amino acid residue interactions of the top 3 compounds screened against Alpha1-6FucT. The drug-residue interactions are color-coded based on the type of interactions.

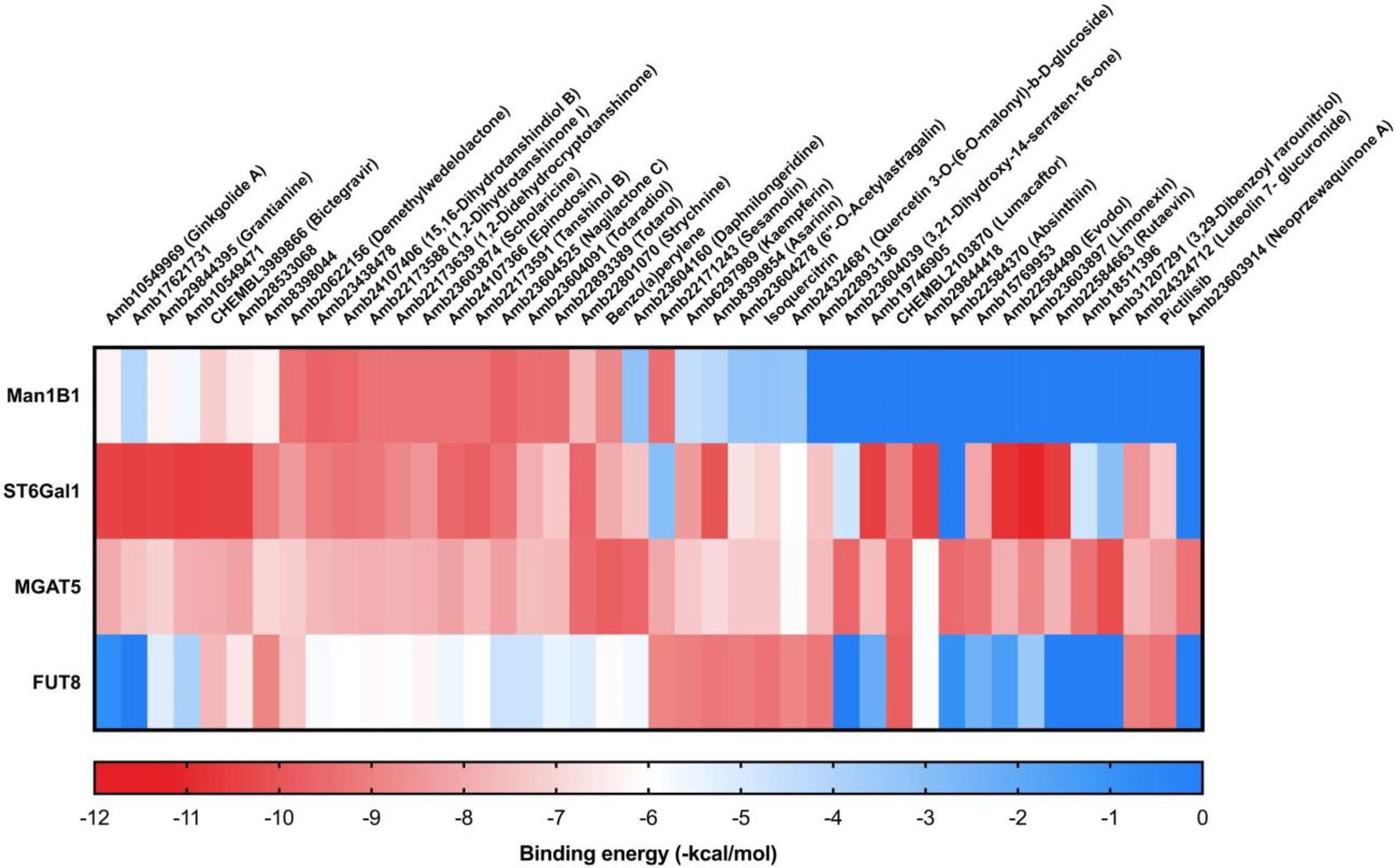


Figure 5. Cross-reactivities (binding to all four enzymes) of the top inhibitors of the four enzymes.

dihydrotanshinone I) [-9.3 kcal/mol], Amb22173591 (tanshinol B) [-9.3 kcal/mol], and Amb22173639 (1,2-didehydrocrypto- tanshinone) [-9.3 kcal/mol] ([Table 3](#_bookmark51)).

## *Alpha1-6FucT screening*

The human Alpha1-6FucT enzyme crystal structure avail- able in databases (PDB ID: 2de0) is in the apo-form. A homology modeling approach was performed using the *C. elegans* Alpha1-6FucT (PDB ID: 3zy6) protein with a bound GDP-fucose. Initially, an automated method was performed using SWISS-MODELLER. However, due to the low sequence homology of both Alpha1-6FucT amino acid sequences, the generated model coverage, as expected, was relatively low ([Supplementary Figure 6](#_bookmark54)). Thus, a chi- meric modeling approach (using SchrodingerTM) was made using the SWISS-MODELLER output and the *C. elegans* Alpha1-6FucT as a template for the human region Alpha1- 6FucT not covered by the SWISS-MODELLER output. After

building the models, the models were superimposed on each other and the corresponding RMSDs were calculated ([Supplementary Figure 7](https://doi.org/10.1080/07391102.2021.2022534)). Kotzler and coworkers (2012) identified several protein regions and amino acid residues with high structural homology between human and *C. ele- gans* Alpha1-6FucT; hence, the RMSD of these regions was calculated (Brzezinski et al., [2012](#_bookmark64)). Based on these compar- isons, the chimeric homology model was the best model obtained.

The screening studies showed that the following compounds gave the highest binding affinity towards the homology-mod- elled Alpha1-6FucT ([Figure 4](#_bookmark48); [Supplementary Figure 8](https://doi.org/10.1080/07391102.2021.2022534)): CHEMBL2103870 (Lumacaftor) [-9.8 kcal/mol], Isoquercitrin [-9.3 kcal/mol], Pictilisib [-9.3 kcal/mol], Amb22893136 [-9.2 kcal/ mol], Amb8399854 (Asarinin) [-9.2 kcal/mol], Amb23604278 (6”- O-acetylastragalin) [-9.1 kcal/mol], Amb24324712 (Luteolin-7- glucuronide) [-9/0 kcal/mol], Amb6297989 (Kaempferin) [-9.0 kcal/mol], Amb22584679 (Sesamolin) [-8.9 kcal/mol], and Amb8398044 [-8.9 kcal/mol].

8 M. R. S. ALVAREZ ET AL.

Table 3. Interactions of key amino acid residues and the top 10 compounds screened against ERMan1.

Binding energy

Ligand

alpha-D-mannopyranose-

(1-2)-methyl 2-thio-alpha- D-mannopyranoside

Amb23438478 Amb23604525

(kcal/mol)

-7.6

Phe329

Pi-sulfur

Glu330

Ile333

Arg334

Hbond

Ala396

Glu397

VDW

Arg461

Hbond

Asp463

Ser464

Leu525

Arg597

Hbond

Pro598

Glu599

Hbond

Phe659

Glu663

Thr688

Hbond

Glu689

Hbond

-9.7

Pi-anion

Pi-anion

Hbond

Pi-alkyl Pi-alkyl

VDW

Hbond Hbond

-9.7

-9.6

Pi-Pi

T shaped Pi-Pi

T shaped Pi-alkyl

Hbond

Pi-alkyl

Amb24107406

Pi-alkyl

Pi-cation

VDW

Pi-sigma

Amb22893389 Amb23604091 Amb20622156

-9.4

Pi-alkyl

Pi-sigma Pi-sigma Pi-alkyl

Pi-sigma Pi-sigma Pi-Pi

T shaped

-9.4

Hbond

Hbond Hbond

-9.3

Hbond

Pi-anion

Hbond

Pi-anion

Amb22173588 Amb22173591

-9.3

Pi-alkyl Pi-alkyl

Pi-anion Pi-anion

Pi-alkyl Pi-alkyl

Pi-alkyl Pi-alkyl

Pi-anion

Hbond Hbond

Pi-alkyl Pi-alkyl

Hbond Hbond

Hbond Hbond

Hbond

Pi-alkyl

Pi-anion

Hbond

Hbond

The binding interactions of the different test ligands with the key enzyme (*Alpha1-6FucT*) active site residues are shown in [Table 4](#_bookmark55). Almost all the ligands interacted with Arg365 via hydrogen bonding while Pictilisib [-9.3 kcal/mol] showed Pi- cation interactions with Arg365. All the top ligands also formed hydrogen bonding interaction with either Ser469 or Gln470. These results are in agreement with a similar dock- ing study by Manabe and coworkers ([2017](#_bookmark86)) where the diphosphate group of GDP-fucose was predicted to form hydrogen bonding interactions with Gly221, Arg365, Ser469, and Gln470. Recent experiments also show that Val450 and

Pi-Pi

T shaped

Pi-sigma

Val471 is important in forming Pi-alkyl interactions with the guanine moiety of GDP (Garc'ıa-Garc'ıa et al., [2020](#_bookmark59)). In our docking experiment, most of the top ligands formed Pi-alkyl interactions with Val450, while some compounds - Amb22893136 [-9.2 kcal/mol], Amb23604278 (6”-O-acetylas- tragalin) [-9.1 kcal/mol], and Amb8398044 [-8.9 kcal/mol] - formed hydrogen bonds with Val471. Additionally, the side- chain residue His363 and backbone residue Tyr250 were shown to tether the guanosine moiety with hydrogen bonds (Garc'ıa-Garc'ıa et al., [2020](#_bookmark59)). In our case, the top ligands CHEMBL2103870 (Lumacaftor) [-9.8 kcal/mol], Isoquercitrin [-9.3 kcal/mol], Pictilisib [-9.3 kcal/mol], Amb23604278 (6”-O- acetylastragalin) [-9.1 kcal/mol], and Amb6297989 (Kaempferin) [-9.0 kcal/mol] formed Pi-Pi T shaped interac- tions with Tyr220 and hydrogen bonding interactions with His363.

Hbond

Hbond

Hbond

Hbond

Hbond

## *Cross-reactivity*

Hbond

By docking the same compound library against multiple enzymes, we evaluated the binding energies of the same compound against several glycosylation enzymes ([Figure 5](#_bookmark49)). Here, the top ligands that were identified against each glyco- sylation enzyme were also compared for their binding affinity against all the other enzymes. Based on the results, most ligands with binding affinity against ST6Gal1 and GlcNAcT-V have low binding affinity against ERMan1. Interestingly, these enzymes catalyze different reactions; GlcNAcT-V and ST6Gal1 add sugar residues to the growing N-glycans during its bio- synthesis, while ERMan1 cleaves off mannose from high-man- nose N-glycans. Furthermore, GlcNAcT-V and ST6Gal1 catalyze reactions that add to hexoses. Also, GlcNAcT-V adds N-acetylglucosamine to mannose while ST6Gal1 adds sialic acid residues to galactose. The distinct reaction catalyzed by the different enzymes could explain the difference in binding affinities between compounds that bind to ST6Gal1 and GlcNAcT-V and Alpha1-6FucT, which adds fucose residues to N-acetylglucosamine.

-9.3

-9.3

-9.3

-9.3

Pi-Pi

T shaped Pi-Pi

T shaped Pi-alkyl

Pi-anion

Pi-alkyl

VDW

Hbond

Pi-alkyl

Hbond

## *ADMETox results*

The ADMETox properties were calculated using the pkCSM software (Pires et al., [2015](#_bookmark88)). The algorithm provides quantita- tive and qualitative predictions of some known ADMETox properties, as well as threshold values. The predicted numer- ical values for Absorption, Distribution, Excretion and Toxicity is shown in [Figure 6](#_bookmark54), with cutoff values implemented by

Amb22173639

Amb23603874

Amb24107366

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS  9

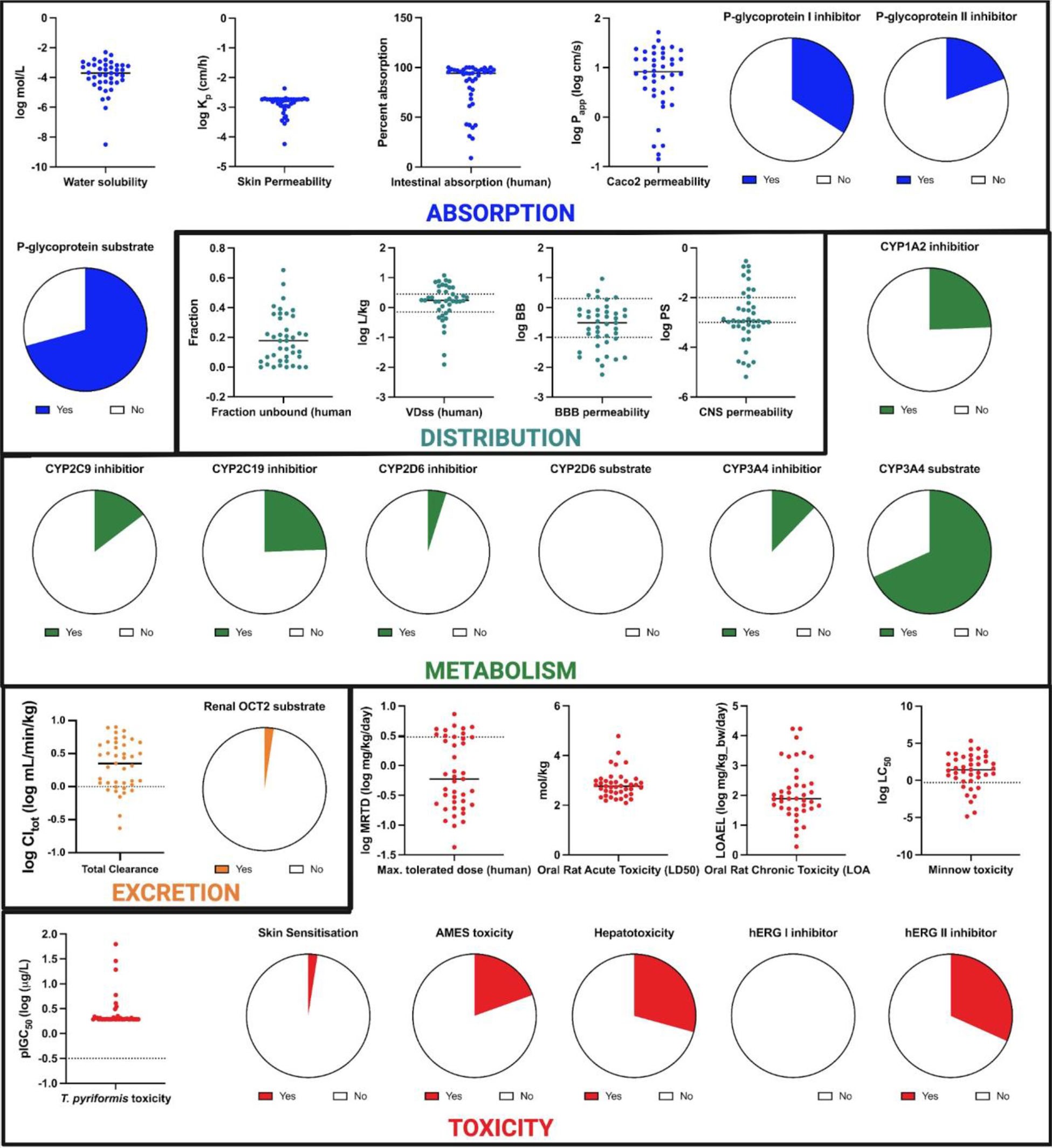


Figure 6. Predicted pharmacokinetic properties of the top inhibitors against the four glycosylation enzymes.

pkCSM represented as dashed lines. For the Absorption parameters, water-solubility, Caco-2 permeability, skin perme- ability, intestinal absorption, P-glycoprotein I/II inhibitor, and P-glycoprotein substrate were predicted. High Caco-2 perme-

ability translates to log Papp > 0.90, while high intestinal absorption translates to %absorption > 30%. P-glycoproteins are ATP-binding cassette (ABC) transporters, functioning as

biological carriers of xenobiotics for extrusion out of cells. Majority of the compounds are considered as P-glycoprotein substrates and non-inhibitors. For Absorption, the following properties were assessed: log P, water-solubility, skin perme- ability, intestinal absorption, and Caco-2 permeability values. Log P values were calculated based on Lipinski’s rule of 5

(Lipinski et al., [2001](#_bookmark80)), which predicts the likelihood of poor absorption of compounds when there are more than 5 H- bond donors, 10 H-bond acceptors, the molecular weight is greater than 500, and Log P greater than five. The predicted

water solubility (Log S) of the compounds at 25 oC is pre-

sented as the logarithm of the molar concentration (log mol/ L). Caco-2 permeability and intestinal absorption values pre- dict absorption of orally administered compounds through the intestines. Compounds with high Caco-2 orally adminis-

tered > 0.90) and high intestinal absorption value (percent- age absorption > 30%) are predicted to have high intestinal

absorption. Compounds that are predicted to have low skin permeability have log Kp values > -2.5. Based on all these

10 M. R. S. ALVAREZ ET AL.

Table 4. Interactions of key amino acid residues and the top 10 compounds screened against Alpha1-6FucT.

Binding

Ligand

Guanosine 50- diphospho-b- L–fucose

CHEMBL2103870

energy

-7.6

Cys218

Hbond

Gly219

Hbond

Tyr220

Hbond

Gly221

Hbond

Cys222

Val258

Phe259

His363

Arg365

Thr367

Hbond

Asp368

Hbond

Thr408

Asp409

Glu444

Asn445

Leu447

Val450

Ser468

Ser469

Gln470

Hbond

Val471

Hbond

Hbond

-9.8

-9.3

-9.3

Pi-Pi

T shaped Pi-Pi

T shaped Pi-Pi

T shaped

Hbond

Hbond

Halogen

(Fluorine)

Pi-alkyl

Halogen (Fluorine)

VDW

Isoquercitrin

Hbond

Pi-Pi

T shaped

Hbond

Hbond

Hbond

Pi-alkyl

Hbond

Hbond

Pictilisib

Pi-alkyl

Hbond

Pi-cation

Hbond

Pi-alkyl

VDW

Amb22893136 Amb8399854

-9.2

Pi-sulfur Pi-Pi

T shaped

Hbond

Hbond Hbond

Hbond VDW

VDW

Pi-alkyl

Pi-alkyl

Hbond Hbond

Hbond

Hbond

-9.2

-9.1

Hbond

Amb23604278

Pi-Pi

T shaped

Hbond

Pi-alkyl

Pi-Pi

T shaped

Hbond

Hbond

Pi-alkyl

Hbond

Hbond

Hbond

Hbond

Hbond

Hbond

Hbond

Hbond Hbond

Hbond

Hbond

Hbond

values, the majority of the top binding ligands were found to have good absorption properties.

In terms of the Distribution parameter, the BBB (blood- brain barrier) and CNS (central nervous system) permeabil- ities, fraction unbound, and VDSs were predicted. The VDSs (steady-state volume of distribution) measure the total dose of a drug that would need to be uniformly distributed to give the same concentration in blood plasma. High VD (log

VDss > 0.45) suggests that a drug is distributed more in tis-

sue rather than in plasma. The BBB (log BB) and CNS (log PS) permeabilities signify the drugs’ ability to penetrate the blood-brain barrier and central nervous system, respectively. For anticancer drugs targeting the lungs, it is desirable for these parameters to be lower: log BB < -1 and log PS < -3. The blood-brain barrier (BBB) and Central Nervous System

Pi-alkyl Pi-alkyl

VDW

(CNS) permeability values predict whether a drug could cross the blood-brain barrier and the central nervous system (through the carotid artery), respectively. For log BBB values

> 0.3 and log PS > -2.0 are considered to have good brain

permeability. Based on the predicted values, less than half of the top binding compounds have good permeability through the blood-brain barrier. The results suggest that these com- pounds may have difficulty accessing the brain; although, low brain permeability values also suggest reduced side effects and off-target effects.

Hbond

Hbond

The Metabolism parameter in ADMETox pertains to com- pounds that are suitable for biotransformations and detoxifi- cations. These enzymatic reactions are catalyzed by cytochrome P450’s such as CYP2D6, CYP3A4, CYP1A2, CYP2C9, and CYP2C19. Therefore, it is desirable for drugs to act as substrates of either of these enzymes and not as inhibitors. Based on the results, majority of the compounds are CYP3A4 substrates.

Hbond Hbond

Pi-Pi

T shaped

Hbond

Hbond

The toxicity parameters were predicted using maximum tolerated dose (MTRD), acute (LD50) and chronic (LOAEL) oral toxicity, T. pyriformis and minnow toxicity, AMES toxicity, and hERG I/II inhibition. MTRD estimates the toxic dose threshold of chemicals in humans. For a given compound, an

Pi-alkyl

MTRD > 0.477 log (mg/kg/day) is considered high. These are

Hbond

the MTRD values obtained in 72 compounds out of all the compounds that were screened. The hERG I and II are genes encoding for potassium channels that are the principal causes of acquired long QT syndrome Our results showed that less than half of the compounds screened are predicted to be hERG I/II inhibitors. The compounds’ toxicities were assessed using Maximum tolerated dose (MRTD), Oral rat acute toxicity (LD50), Oral rat chronic toxicity (LOAE), and

Pi-Pi

T shaped Pi-Pi

T shaped

Minnow toxicity values. For MRTD, values < 0.477 log (mg/

kg/day) are considered low, whereas for minnow toxicity, log LC50 values<-0.3 are considered high. Based on the cumula- tive scores from ADMETox, all the top compounds screened have good drug-likeness ([Supplementary Table 3](#_bookmark51)).

-9.0

Pi-alkyl Hbond

-9.0

-8.9

-8.9

Hbond

# Conclusion

Amb24324712 Amb6297989

Amb22584679

Amb8398044

We employed a high throughput *in silico* docking method- ology to predict small molecules that can inhibit four glyco- sylation enzymes that are associated with cancer

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS 11

progression. By comparing the predicted binding affinities of the small molecules against the four target glycosylation enzymes, we identified several small molecules that could bind to more than one glycosylation enzyme, paving the way for future development of multi-targeting glycosylation inhibitors. Furthermore, we also assessed the pharmacoki- netic properties of our top inhibitors and have found them to be favorable. Currently, efforts are underway in perform- ing *in vitro* and *in vivo* mass spectrometry-based assays to assess the efficacies of some of these compounds in inhibit- ing glycosylation in lung cancer cell models.

# Disclosure statement

No potential conflict of interest was reported by the author(s).

# References

Brzezinski, K., Dauter, Z., & Jaskolski, M. [(2012).](#_bookmark50) Structures of NodZ a1,6- fucosyltransferase in complex with GDP and GDP-fucose. *Acta Crystallographica. Section D, Biological Crystallography*, *68*(Pt 2), 160–168. <https://doi.org/10.1107/S0907444911053157>

Burchell, J., Poulsom, R., Hanby, A., Whitehouse, C., Cooper, L., Clausen, H., Miles, D., & Taylor-Papadimitriou, J. [(1999).](#_bookmark10) An alpha2,3 sialyltrans- ferase (ST3Gal I) is elevated in primary breast carcinomas. *Glycobiology*, *9*(12), 1307–1311. [https://doi.org/10.1093/glycob/9.12.](https://doi.org/10.1093/glycob/9.12.1307)

[1307](https://doi.org/10.1093/glycob/9.12.1307)

Chen, X., Wu, J., Huang, H., Ding, Q., Liu, X., Chen, L., Zha, X., Liang, M., He, J., Zhu, Q., Wang, S., & Xia, T. [(2016).](#_bookmark11) Comparative profiling of tri- ple-negative breast carcinomas tissue glycoproteome by sequential purification of glycoproteins and stable isotope labeling. *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology*, *38*(1), 110–121. <https://doi.org/10.1159/000438613>

Choudhary, N., & Singh, V. [(2019).](#_bookmark4) Insights about multi-targeting and synergistic neuromodulators in Ayurvedic herbs against epilepsy: inte- grated computational studies on drug-target and protein-protein interaction networks. *Scientific Reports*, *9*(1), 10565. [https://doi.org/10.](https://doi.org/10.1038/s41598-019-46715-6) [1038/s41598-019-46715-6](https://doi.org/10.1038/s41598-019-46715-6)

Cohen, E. N., Fouad, T. M., Lee, B.-N., Arun, B. K., Liu, D., Tin, S., Gutierrez Barrera, A. M., Miura, T., Kiyokawa, I., Yamashita, J., Alvarez, R. H., Valero, V., Woodward, W. A., Shen, Y., Ueno, N. T., Cristofanilli, M., & Reuben, J. M. [(2019).](#_bookmark11) Elevated serum levels of sialyl Lewis X (sLeX) and inflammatory mediators in patients with breast cancer. *Breast Cancer Research and Treatment*, *176*(3), 545–556. [https://doi.org/10.](https://doi.org/10.1007/s10549-019-05258-0) [1007/s10549-019-05258-0](https://doi.org/10.1007/s10549-019-05258-0)

Coria-Te'llez, A. V., Montalvo-Go'nzalez, E., Yahia, E. M., & Obledo-V'azquez,

E. N. [(2018).](#_bookmark22) Annona muricata: A comprehensive review on its trad- itional medicinal uses, phytochemicals, pharmacological activities, mechanisms of action and toxicity. *Arabian Journal of Chemistry*, *11*(5), 662–691. <https://doi.org/10.1016/j.arabjc.2016.01.004>

D[Zbreve]Ami'c, A. M., Marin, P. D., Gbolade, A. A., & Risti'c, M. S. [(2010).](#_bookmark23)

Chemical composition of Mangifera indica essential oil from Nigeria. *Journal of Essential Oil Research*, *22*(2), 123–125. [https://doi.org/10.](https://doi.org/10.1080/10412905.2010.9700279) [1080/10412905.2010.9700279](https://doi.org/10.1080/10412905.2010.9700279)

Dallakyan, S., & Olson, A. J. [(2015).](#_bookmark24) *Small-molecule library screening by docking with PyRx BT - Chemical biology: Methods and protocols* (J. E. Hempel, C. H. Williams, & C. C. Hong, eds.). New York, NY: Humana Press. <https://doi.org/10.1007/978-1-4939-2269-7_19>

Dall’Olio, F., & Chiricolo, M. [(2001).](#_bookmark12) Sialyltransferases in cancer. *Glycoconjugate Journal*, *18*(11/12), 841–850. https://doi.org/10.1023/ A:1022288022969

Dall’Olio, F., Malagolini, N., Trinchera, M., & Chiricolo, M. [(2012).](#_bookmark13) Mechanisms of cancer-associated glycosylation changes. *Frontiers in Bioscien*ce. (Landmark Edition), *17*, 670–699. [https://doi.org/10.2741/](https://doi.org/10.2741/3951)

[3951](https://doi.org/10.2741/3951)

Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., Wiegers, J., Wiegers,

T. C., & Mattingly, C. J. [(2021).](#_bookmark43) Comparative Toxicogenomics Database (CTD): update 2021. *Nucleic Acids Research*, *49*(D1), D1138–D1143. <https://doi.org/10.1093/nar/gkaa891>

Egan, W. J., Merz, K. M., & Baldwin, J. J. [(2000).](#_bookmark40) Prediction of drug absorp- tion using multivariate statistics. *Journal of Medicinal Chemistry*, *43*(21), 3867–3877. <https://doi.org/10.1021/jm000292e>

Freshour, S. L., Kiwala, S., Cotto, K. C., Coffman, A. C., McMichael, J. F.,

Song, J. J., Griffith, M., Griffith, O. L., & Wagner, A. H. [(2021).](#_bookmark43) Integration of the drug-gene interaction database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Research*, *49*(D1), D1144–D1151. <https://doi.org/10.1093/nar/gkaa1084>

Garc'ıa-Garc'ıa, A., Ceballos-Laita, L., Serna, S., Artschwager, R., Reichardt,

N. C., Corzana, F., & Hurtado-Guerrero, R. [(2020).](#_bookmark53) Structural basis for substrate specificity and catalysis of a1,6-fucosyltransferase. *Nature Communications*, *11*(1), 973. [https://doi.org/10.1038/s41467-020-](https://doi.org/10.1038/s41467-020-14794-z) [14794-z](https://doi.org/10.1038/s41467-020-14794-z)

Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. [(1999).](#_bookmark41) A know- ledge-based approach in designing combinatorial or medicinal chem- istry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *Journal of Combinatorial Chemistry*, *1*(1), 55–68. <https://doi.org/10.1021/cc9800071>

Green, D. V. S. [(2003).](#_bookmark5) Virtual screening of virtual libraries. *Progress in Medicinal Chemistry*, *41*, 61–97. [https://doi.org/10.1016/s0079-](https://doi.org/10.1016/s0079-6468(02)41002-8)

[6468(02)41002-8](https://doi.org/10.1016/s0079-6468(02)41002-8)

H€ahnke, V. D., Kim, S., & Bolton, E. E. [(2018).](#_bookmark6) PubChem chemical structure standardization. *Journal of Cheminformatics*, *10*(1), 36. [https://doi.org/](https://doi.org/10.1186/s13321-018-0293-8) [10.1186/s13321-018-0293-8](https://doi.org/10.1186/s13321-018-0293-8)

Handerson, T., Camp, R., Harigopal, M., Rimm, D., & Pawelek, J. [(2005).](#_bookmark12) Beta1,6-branched oligosaccharides are increased in lymph node meta- stases and predict poor outcome in breast carcinoma. *Clinical Cancer Research*, *11*(8), 2969–2973. [https://doi.org/10.1158/1078-0432.CCR-04-](https://doi.org/10.1158/1078-0432.CCR-04-2211)

[2211](https://doi.org/10.1158/1078-0432.CCR-04-2211)

Ihara, H., Ikeda, Y., Toma, S., Wang, X., Suzuki, T., Gu, J., Miyoshi, E., Tsukihara, T., Honke, K., Matsumoto, A., Nakagawa, A., & Taniguchi, N. [(2007).](#_bookmark32) Crystal structure of mammalian alpha1,6-fucosyltransferase, FUT8. *Glycobiology*, *17*(5), 455–466. [https://doi.org/10.1093/glycob/](https://doi.org/10.1093/glycob/cwl079) [cwl079](https://doi.org/10.1093/glycob/cwl079)

Kannagi, R., Yin, J., Miyazaki, K., & Izawa, M. [(2008).](#_bookmark14) Current relevance of incomplete synthesis and neo-synthesis for cancer-associated alter- ation of carbohydrate determinants-Hakomori’s concepts revisited. *Biochimica et Biophysica Acta*, *1780*(3), 525–531. [https://doi.org/10.](https://doi.org/10.1016/j.bbagen.2007.10.007) [1016/j.bbagen.2007.10.007](https://doi.org/10.1016/j.bbagen.2007.10.007)

Kapetanovic, I. M. [(2008).](#_bookmark7) Computer-aided drug discovery and develop- ment (CADDD): in silico-chemico-biological approach. *Chemico- Biological Interactions*, *171*(2), 165–176. [https://doi.org/10.1016/j.cbi.](https://doi.org/10.1016/j.cbi.2006.12.006)

[2006.12.006](https://doi.org/10.1016/j.cbi.2006.12.006)

Karaveg, K., Siriwardena, A., Tempel, W., Liu, Z.-J., Glushka, J., Wang, B.-C., & Moremen, K. W. [(2005).](#_bookmark33) Mechanism of class 1 (glycosylhydrolase family 47) alpha -mannosidases involved in N-glycan processing and endoplasmic reticulum quality control. *The Journal of Biological Chemistry*, *280*(16), 16197–16207. [https://doi.org/10.1074/jbc.](https://doi.org/10.1074/jbc.M500119200) [M500119200](https://doi.org/10.1074/jbc.M500119200)

f g

Kuhn, B., Benz, J., Greif, M., Engel, A. M., Sobek, H., & Rudolph, M. G. [(2013).](#_bookmark34) The structure of human a-2,6-sialyltransferase reveals the binding mode of complex glycans. *Acta Crystallographica. Section D, Biological Crystallography*, *69*(Pt 9), 1826–1838. [https://doi.org/10.](https://doi.org/10.1107/S0907444913015412) [1107/S0907444913015412](https://doi.org/10.1107/S0907444913015412)

Landi, M. T., Dracheva, T., Rotunno, M., Figueroa, J. D., Liu, H., Dasgupta, A., Mann, F. E., Fukuoka, J., Hames, M., Bergen, A. W., Murphy, S. E., Yang, P., Pesatori, A. C., Consonni, D., Bertazzi, P. A., Wacholder, S., Shih, J. H., Caporaso, N. E., & Jen, J. [(2008).](#_bookmark12) Gene expression signature of cigarette smoking and its role in lung adenocarcinoma develop- ment and survival. *PLOS One*, *3*(2), e1651. [https://doi.org/10.1371/jour-](https://doi.org/10.1371/journal.pone.0001651) [nal.pone.0001651](https://doi.org/10.1371/journal.pone.0001651)

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. [(2001).](#_bookmark41) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, *46*(1–3), 3–26. [https://doi.org/10.](https://doi.org/10.1016/S0169-409X(00)00129-0) [1016/S0169-409X(00)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0)

12 M. R. S. ALVAREZ ET AL.

Lira-Navarrete, E., Valero-Gonz'alez, J., Villanueva, R., Mart'ınez-Ju'lvez, M., Tejero, T., Merino, P., Panjikar, S., & Hurtado-Guerrero, R. [(2011).](#_bookmark35) Structural insights into the mechanism of protein O-Fucosylation. *PLOS One*, *6*(9), e25365. <https://doi.org/10.1371/journal.pone.0025365>

Liu, Y.-C., Yen, H.-Y., Chen, C.-Y., Chen, C.-H., Cheng, P.-F., Juan, Y.-H.,

Chen, C.-H., Khoo, K.-H., Yu, C.-J., Yang, P.-C., Hsu, T.-L., & Wong, C.-H.

[(2011).](#_bookmark15) Sialylation and fucosylation of epidermal growth factor recep- tor suppress its dimerization and activation in lung cancer cells. *Proceedings of the National Academy of Sciences of the United States of* *America*, *108*(28), 11332–11337. [https://doi.org/10.1073/pnas.](https://doi.org/10.1073/pnas.1107385108)

[1107385108](https://doi.org/10.1073/pnas.1107385108)

Manabe, Y., Kasahara, S., Takakura, Y., Yang, X., Takamatsu, S., Kamada, Y., Miyoshi, E., Yoshidome, D., & Fukase, K. [(2017).](#_bookmark52) Development of a1,6-fucosyltransferase inhibitors through the diversity-oriented syn- theses of GDP-fucose mimics using the coupling between alkyne and sulfonyl azide. *Bioorganic & Medicinal Chemistry*, *25*(11), 2844–2850. <https://doi.org/10.1016/j.bmc.2017.02.036>

Manosroi, A., Jantrawut, P., Sainakham, M., Manosroi, W., & Manosroi, J. [(2012).](#_bookmark25) AAnticanceractivities of the extract from Longkong (Lansium domesticum) young fruits. *Pharmaceutical Biology*, *50*(11), 1397–1407. <https://doi.org/10.3109/13880209.2012.682116>

Marfori, E., Kajiyama, S., Fukusaki, E., & Kobayashi, A. [(2015).](#_bookmark26) Lansioside D, a new triterpenoid glycoside antibiotic from the fruit peel of Lansium domesticum Correa. *Journal of Pharmacognosy and* *Phytochemistry*, *3*, 140–143.

Martin, Y. C. [(2005).](#_bookmark42) A bioavailability score. *Journal of Medicinal* *Chemistry*, *48*(9), 3164–3170. <https://doi.org/10.1021/jm0492002>

Masibo, M., & He, Q. [(2008).](#_bookmark23) Major mango polyphenols and their poten- tial significance to human health. *Comprehensive Reviews in Food Science and Food Safety*, *7*(4), 309–319. [https://doi.org/10.1111/j.1541-](https://doi.org/10.1111/j.1541-4337.2008.00047.x) [4337.2008.00047.x](https://doi.org/10.1111/j.1541-4337.2008.00047.x)

Meany, D. L., & Chan, D. W. [(2011).](#_bookmark27) Aberrant glycosylation associated with enzymes as cancer biomarkers. *Clinical Proteomics*, *8*(1), 7. <https://doi.org/10.1186/1559-0275-8-7>

Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., F'elix, E.,

Magarin~os, M. P., Mosquera, J. F., Mutowo, P., Nowotka, M., Gordillo- Maran~o'n, M., Hunter, F., Junco, L., Mugumbate, G., Rodriguez-Lopez, M., Atkinson, F., Bosc, N., Radoux, C. J., Segura-Cabrera, A., Hersey, A., & Leach, A. R. [(2019).](#_bookmark8) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Research*, *47*(D1), D930–D940. [https://doi.org/10.](https://doi.org/10.1093/nar/gky1075) [1093/nar/gky1075](https://doi.org/10.1093/nar/gky1075)

Mohs, R. C., & Greig, N. H. [(2017).](#_bookmark3) Drug discovery and development: Role of basic biological research. *Alzheimer*’*s & Dementia (New York, NY)*, *3*(4), 651–657. <https://doi.org/10.1016/j.trci.2017.10.005>

Muegge, I., Heald, S. L., & Brittelli, D. [(2001).](#_bookmark42) Simple selection criteria for drug-like chemical matter. *Journal of Medicinal Chemistry*, *44*(12), 1841–1846. <https://doi.org/10.1021/jm015507e>

Munkley, J. [(2016).](#_bookmark16) The role of Sialyl-Tn in cancer. *International Journal of**Molecular Sciences*, *17*(3), 275. <https://doi.org/10.3390/ijms17030275> Nagae, M., Kizuka, Y., Mihara, E., Kitago, Y., Hanashima, S., Ito, Y., Takagi,

J., Taniguchi, N., & Yamaguchi, Y. [(2018).](#_bookmark31) Structure and mechanism of cancer-associated N-acetylglucosaminyltransferase-V. *Nature Communications*, *9*(1), 3380. [https://doi.org/10.1038/s41467-018-](https://doi.org/10.1038/s41467-018-05931-w) [05931-w](https://doi.org/10.1038/s41467-018-05931-w)

Nu'n~ez Sell'es, A. J., V'elez Castro, H. T., Agu€ero-Agu€ero, J., Gonz'alez-

Gonz'alez, J., Naddeo, F., De Simone, F., & Rastrelli, L. [(2002).](#_bookmark23) Isolation and quantitative analysis of phenolic antioxidants, free sugars, and polyols from mango (Mangifera indica L.) stem bark aqueous decoc- tion used in Cuba as a nutritional supplement. *Journal of Agricultural* *and Food Chemistry*, *50*(4), 762–766. <https://doi.org/10.1021/jf011064b>

O’Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., &

Hutchison, G. R. [(2011).](#_bookmark28) Open babel: An open chemical toolbox.

*Journal of Cheminformatics*, *3*(1), 33. [https://doi.org/10.1186/1758-](https://doi.org/10.1186/1758-2946-3-33)

[2946-3-33](https://doi.org/10.1186/1758-2946-3-33)

Pence, H. E., & Williams, A. [(2010).](#_bookmark6) ChemSpider: An online chemical infor- mation resource. *Journal of Chemical Education*, *87*(11), 1123–1124. <https://doi.org/10.1021/ed100697w>

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt,

D. M., Meng, E. C., & Ferrin, T. E. [(2004).](#_bookmark36) UCSF Chimera-a visualization

system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>

Picco, G., Julien, S., Brockhausen, I., Beatson, R., Antonopoulos, A., Haslam, S., Mandel, U., Dell, A., Pinder, S., Taylor-Papadimitriou, J., & Burchell, J. [(2010).](#_bookmark15) Over-expression of ST3Gal-I promotes mammary tumorigenesis. *Glycobiology*, *20*(10), 1241–1250. [https://doi.org/10.](https://doi.org/10.1093/glycob/cwq085) [1093/glycob/cwq085](https://doi.org/10.1093/glycob/cwq085)

Pinho, S. S., & Reis, C. A. [(2015).](#_bookmark15) Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews. Cancer*, *15*(9), 540–555. [https://](https://doi.org/10.1038/nrc3982) [doi.org/10.1038/nrc3982](https://doi.org/10.1038/nrc3982)

Pino, J. A., & Mesa, J. [(2006).](#_bookmark22) Contribution of volatile compounds to mango (Mangifera indica L.) aroma. *Flavour and Fragrance Journal*, *21*(2), 207–213. <https://doi.org/10.1002/ffj.1703>

Pires, D. E. V., Blundell, T. L., & Ascher, D. B. [(2015).](#_bookmark29) pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph- based signatures. *Journal of Medicinal Chemistry*, *58*(9), 4066–4072. <https://doi.org/10.1021/acs.jmedchem.5b00104>

Potapenko, I. O., Haakensen, V. D., Lu€ders, T., Helland, A., Bukholm, I.,

Sørlie, T., Kristensen, V. N., Lingjaerde, O. C., & Børresen-Dale, A.-L. [(2010).](#_bookmark17) Glycan gene expression signatures in normal and malignant breast tissue; possible role in diagnosis and progression. *Molecular Oncology*, *4*(2), 98–118. <https://doi.org/10.1016/j.molonc.2009.12.001>

Pozzan, A. [(2006).](#_bookmark9) Molecular descriptors and methods for ligand based virtual high throughput screening in drug discovery. *Current Pharmaceutical Design*, *12*(17), 2099–2110. [https://doi.org/10.2174/](https://doi.org/10.2174/138161206777585247)

[138161206777585247](https://doi.org/10.2174/138161206777585247)

Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A., & Skiff, W. M. [(1992).](#_bookmark28) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, *114*(25), 10024–10035. [https://doi.org/10.1021/](https://doi.org/10.1021/ja00051a040) [ja00051a040](https://doi.org/10.1021/ja00051a040)

Recchi, M. A., Hebbar, M., Hornez, L., Harduin-Lepers, A., Peyrat, J. P., & Delannoy, P. [(1998).](#_bookmark17) Multiplex reverse transcription polymerase chain reaction assessment of sialyltransferase expression in human breast cancer. *Cancer Research*, *58*(18), 4066–4070.

Reily, C., Stewart, T. J., Renfrow, M. B., & Novak, J. [(2019).](#_bookmark18) Glycosylation in health and disease. *Nature Reviews. Nephrology*, *15*(6), 346–366. <https://doi.org/10.1038/s41581-019-0129-4>

Ruhaak, L. R., Taylor, S. L., Stroble, C., Nguyen, U. T., Parker, E. A., Song,

T., Lebrilla, C. B., Rom, W. N., Pass, H., Kim, K., Kelly, K., & Miyamoto, S. [(2015).](#_bookmark17) Differential N-glycosylation patterns in lung adenocarcinoma tissue. *Journal of Proteome Research*, *14*(11), 4538–4549. [https://doi.](https://doi.org/10.1021/acs.jproteome.5b00255) [org/10.1021/acs.jproteome.5b00255](https://doi.org/10.1021/acs.jproteome.5b00255)

Ruhaak, L. R., Xu, G., Li, Q., Goonatilleke, E., & Lebrilla, C. B. [(2018).](#_bookmark19) Mass spectrometry approaches to glycomic and glycoproteomic analyses. *Chemical Reviews*, *118*(17), 7886–7930. [https://doi.org/10.1021/acs.](https://doi.org/10.1021/acs.chemrev.7b00732) [chemrev.7b00732](https://doi.org/10.1021/acs.chemrev.7b00732)

Sorokina, M., & Steinbeck, C. [(2020).](#_bookmark4) Review on natural products data- bases: where to find data in 2020. *Journal of Cheminformatics*, *12*(1), 20. <https://doi.org/10.1186/s13321-020-00424-9>

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T. I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., & Lancet, D. [(2016).](#_bookmark44) The GeneCards Suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics*, *54*, 1.30.1–1.30.33. <https://doi.org/10.1002/cpbi.5>

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. v. [(2019).](#_bookmark44) STRING v11: protein-protein association net- works with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. <https://doi.org/10.1093/nar/gky1131>

Trott, O., & Olson, A. J. [(2010).](#_bookmark38) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimiza- tion, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. <https://doi.org/10.1002/jcc.21334>

Veber, D. F., Johnson, S. R., Cheng, H.-Y., Smith, B. R., Ward, K. W., & Kopple, K. D. [(2002).](#_bookmark40) Molecular properties that influence the oral bio- availability of drug candidates. *Journal of Medicinal Chemistry*, *45*(12), 2615–2623. <https://doi.org/10.1021/jm020017n>

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS 13

Venkitachalam, S., & Guda, K. [(2017).](#_bookmark20) Altered glycosyltransferases in colo- rectal cancer. *Expert Review of Gastroenterology & Hepatology*, *11*(1), 5–7. <https://doi.org/10.1080/17474124.2017.1253474>

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. [(2018).](#_bookmark37) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. <https://doi.org/10.1093/nar/gky427>

Weinstein, J., Lee, E. U., McEntee, K., Lai, P. H., & Paulson, J. C. [(1987).](#_bookmark21) Primary structure of beta-galactoside alpha 2,6-sialyltransferase. Conversion of membrane-bound enzyme to soluble forms by cleav- age of the NH2-terminal signal anchor. *The Journal of Biological*

*Chemistry*, *262*(36), 17735–17743. [https://doi.org/10.1016/S0021-](https://doi.org/10.1016/S0021-9258(18)45441-5)

[9258(18)45441-5](https://doi.org/10.1016/S0021-9258(18)45441-5)

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., … Wilson, M. [(2018).](#_bookmark44) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, *46*(D1), D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>

Wouters, O. J., McKee, M., & Luyten, J. [(2020).](#_bookmark3) Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA*, *323*(9), 844–853. [https://doi.org/10.1001/jama.2020.](https://doi.org/10.1001/jama.2020.1166)

[1166](https://doi.org/10.1001/jama.2020.1166)