A Translational Drug Screening Platform for Improving the Safety of First-Line Anti-Cancer Therapeutics

By

Anup Prasant Aron Challa

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Chemical Engineering

May 14, 2021

Nashville, Tennessee

Approved:

Ethan S. Lippmann, PhD

G. Kane Jennings, PhD

ACKNOWLEDGEMENTS

The Intramural Research Program of the National Center for Advancing Translational Sciences (NCATS) of the National Insitutes of Health (NIH) provided financial and infrastructural support for the work I present in this thesis. I also appreciate support from Clinical and Translational Science Award U54TR02243-02 from NCATS and acknowledge that the content of this thesis does not represent the official views of NIH or Symberix, Inc., the industrial co-sponsor of the work I present here.

This research is a success I share with several of my talented co-investigators at NCATS, including Min Shen, PhD, Informatics Group Leader, whom I thank profusely for her dedicated and continuing advocacy for my professional development; Xin Hu, PhD, Informatician; Ya-Qin Zhang, PhD, Biologist; Surendra Karavadhi, PhD, Chemist; Samarjit Patnaik, PhD, Chemistry Group Leader; and Matthew Hall, PhD, Branch Chief. At Symberix, Bret Wallace, PhD, Director of Research, and Jeffery Hymes, PhD, Scientist, also contributed to defining the framework of the investigation I present here by providing stimulus materials for our research. Drs. Zhang and Hymes completed all the bench work that I discuss in this thesis as validative of the informatics models and hit prioritization approaches that I developed.

I am grateful for the supportive environment that the Department of Chemical and Biomolecular Engineering and the Vanderbilt Institute for Clinical and Translational Research have provided me for expansive, self-driven inquiry throughout my time in the Accelerated Graduate Program in Engineering. I also appreciate the faculty of the Department of Biomedical

Informatics at Harvard Medical School for naming me Visiting Research Fellow during my time in the Program and for the experiences this position has allowed me embrace.

I thank immensely Kane Jennings, PhD, Professor and Chair of the Department of Chemical and Biomolecular Engineering, for his patience in helping me to meaningfully navigate the non-linearity of the Accelerated Graduate Program in Engineering and for his ever-thoughtful mentorship.

I am extremely grateful for the support of Ethan Lippmann, PhD, Assistant Professor of the Department of Chemical and Biomolecular Engineering, and David Aronoff, MD, FIDSA, FAAM, Professor and Addison B. Scoville, Jr. Chair of the Department of Medicine, for co-advising this thesis and for serving simultaneously as my colleagues and mentors. Their dedicated evangelism towards my success and their receptiveness towards mutual, non-traditional collaboration have provided me with singularly impactful training in translational science. Their lessons in servant leadership continue to shape my professional values, and I am excited for the team science we have on the horizon.

TABLE OF CONTENTS

Page

Appendix

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

Term	Abbreviation
(human/microbial) β-glucuronidase enzymes	(h/b)GUS
amino acid	AA
Basic Local Alignment Search Tool	BLAST
curve class	CC
drug metabolism and pharmacokinetics	DMPK
gastrointestinal	GI
(virtual) high-throughput screening	(v)HTS
Iterative Threading ASSEmbly Refinement	I-TASSER
mechanism of action	MOA
Molecular Operating Environment	MOE
National Center for Advancing Translational Sciences	NCATS
National Chemical Genomics Center	NCGC
National Institutes of Health	NIH
non-steroidal, anti-inflammatory drug	NSAID
point accepted mutation	PAM
Protein Data Bank	PDB
RUC-4 inhibitors	RUC-4i

Sequence Editor	SE
serious, adverse event	SAE
signal-to-background ratio	S:B
standard-of-care	POC
template modeling score	TM
user interface	UI
uridine 5’-diphospho-glucuronosyltransferase	UGT

CHAPTER I

INTRODUCTION

In the age of data-driven medicine, drug development benefits from analysis of post-market surveillance data, such that problematic use of existing drugs may inspire new design strategies to optimize the use of these therapeutics[1–3]. This is especially relevant in cancer, a spectrum of diseases for which pharmacological intervention may be within the first line of therapeutic remediation. Most anti-cancer treatment modalities act through promoting death of rapidly dividing cells, such that they may suppress tumorigenesis but can also generate significant "off-target" effects from loss of epithelial cell volume across a patient's organ systems[4–6]. Phenotypes emerging from this widespread apoptosis may include nausea and vomiting, loss of appetite, hair loss, and the potential for development of an immunocompromised state, given the high replication rates of innate and adaptive immune cells[7–9]. Therefore, non-adherence to anti-cancer treatment remains high among neoplastic patients, as patients seek to avoid severe side effects of their drug regimens[10–12]. This problem speaks to the ongoing need for new therapeutic strategies that may allow for suppression of these "off-target" effects, consequently improving the therapeutic index of anti-cancer drugs by reducing dose-limiting toxicity.

Enteric processing of anti-cancer drugs to urine-excretable metabolites forms non-toxic drug-glucuronides by uridine 5'-diphospho-glucuronosyltransferase (UGT)-mediated metabolism in the liver. In turn, β-glucuronidase enzymes (GUS) among commensal gut microbiota act on the drug-glucuronides to cleave their sugar moieties for nutrition, releasing

reactivated drug molecules in the presence of vulnerable enterocytes. Consequently, the microbiome-dependent reactivation of small-molecule anti-cancer drugs in the presence of vulnerable enterocytes may be responsible for gastrointestinal (GI) serious, adverse events (SAEs), reducing patient adherence to their anti-cancer treatment regimens[13–15]. Indeed, this pattern of drug metabolism and pharmacokinetics (DMPK) is associated with at least 279 unique strains of GUS within the diverse microbial flora of the human gut[16]. I note that though humans also contain an endemic analog of GUS that is structurally heterogeneous to those in their microbiomes, the high density of bacterial GUS analogs renders the human analog minimally active, compared to that within resident gut bacteria[17]. For the remainder of this article, I will distinguish bacterial GUS and human GUS analogs as *bGUS* and *hGUS*, respectively.

As Ervin *et al*. discuss, a bGUS-inhibitory prodrug may increase the safety and tolerability of first-line anti-cancer agents: these investigators unpacked the metabolic fate of the anti-neoplastic agent regorafenib in the liver and GI tract[18]. In turn, they found that raloxifene, a drug approved by United States Food and Drug Administration for the treatment and prevention of osteoporosis in post-menopausal women[19], is a potential inhibitor of the bGUS enzymes that re-activate regorafenib in the GI tract. Herein, the results of this study suggest that delivering a therapeutic adjunct capable of bGUS blockade could potentially reduce the burden of patient SAEs associated with small-molecule anti-cancer therapy[20].

The mechanism of action (MOA) of bGUS in cleaving glucuronic acid moieties from glucuronidated compounds is well-described within the relevant biochemical literature and resembles that classically implicated in peptic ulceration upon excessive patient exposure to non-steroidal, anti-inflammatory drugs (NSAIDs)[21]. Therefore, given that this article does not seek to probe the mechanistic details of the bGUS enzyme's function, I refer readers with an interest in

deep discussion of this information to work of Wallace *et al*., Ervin *et al*., and Awolade *et al*., who present a thorough description of bGUS MOA[18,22,23]. These authors also describe the protein biochemistry that underlies the nature of modulatory effects on this class of enzymes; I will present the relevance and validation of this information—within the scope of the present study—in an *ad hoc* fashion throughout my thesis, as I focus on the application of *in silico* high-throughput screening (HTS) as a method of novel and previously-unattempted discovery for precision inhibitors of homologous bGUS analogs, rather than on the granularity of the established biochemistry that underlies these interactions. Indeed, these interactions are critical to understanding the relevance of our efforts to GI physiology but are most relevant to this study as validative information for our models.

Orthologs of bGUS exhibit a wide range of catalytic efficiencies for glucuronide substrates, driven by key structural differences among homologous enzymes[16,24–26]. Despite this structural and functional heterogeneity, previous drug screening efforts have focused solely on characterizing the bGUS ortholog encoded by gut commensal *E. coli*. This is a major weakness of the existing literature because most bGUS orthologs share < 25% amino acid identity with *E. coli* GUS[16].

Therefore, the goal of the present study is to develop a robust HTS platform that can identify inhibitors of common bGUS analogs in the human microbiome. While broad-spectrum antibiotics, such as cephalosporins, fluoroquinolones, and tetracyclines, offer the ability to suppress microbial action on glucuronidated drugs, these drugs can also suppress the beneficial activity of the gut microbiome[27–29]. Therefore, since such modulation could generate SAEs similar to—or potentially in augmentation to—those resulting from exposure to anti-cancer therapies[30], a precision inhibition strategy that affects only bGUS, while maintaining the

beneficial activity of the human microbiome, is critical to the development of a drug within this space. This necessary precision specifies a difficult drug design task, which requires consideration of a very large set of potential inhibitors to identify agents with the most optimal activity profiles.

Virtual high-throughput drug screening (vHTS) has emerged as an efficient method of targeted drug discovery, as it allows for simulation of a drug's mode of binding to a specified target and thereby generates binding energy parameters that may act as predictive hypotheses of efficacy[31–33]. When the structure of a target is available or feasibly predictable, this technique is an efficient method of prioritizing compounds for further interrogation as potential hits: though a standard HTS approach allows for fast, combinatorial screening of compounds of interest towards a specific target, the combinatorial dimensionality of this technique results in significant material and labor costs and has low efficiency[34]. Therefore, in contrast to the ~1% hit rate that most HTS panels achieve, vHTS—with follow-up testing *in vitro* of only compounds with the highest predicted activity—can achieve nearly an order of magnitude more success, as evidenced by hit rates ~3-4% and significantly less material costs in the interrogation of promising hits[35]. Figure 1 below presents the limiting tradeoff of scale and discovery inherent to HTS, which vHTS seeks to bridge.

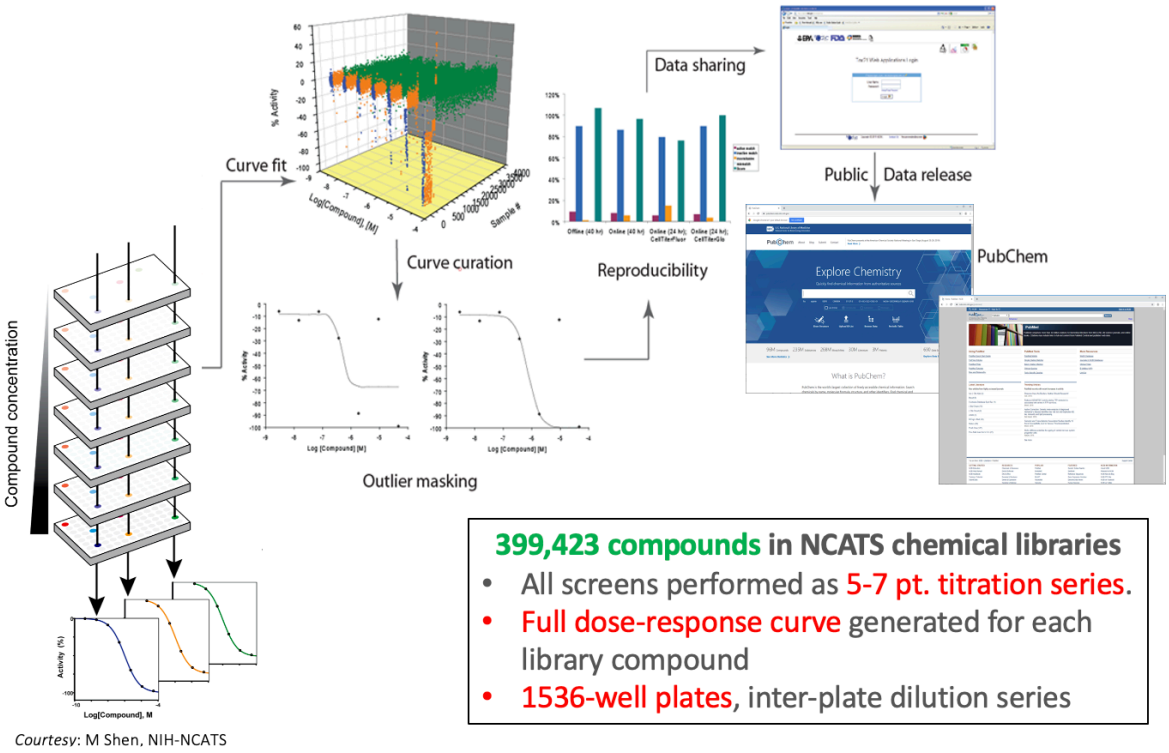

Figure 1: While HTS can promote of new hits via fully agnostic screening, it is resource-intensive and, attributable to its expansive scale, results in low hit discovery rates.

Therefore, given the well-defined target rationale of our drug design task—as well as the challenge of establishing precision blockade of bGUS—we established a first-in-kind vHTS platform to discover inhibitors of bGUS and identify hits towards the development of a potential therapeutic adjunct to reduce the frequency of SAEs associated with first-line anti-cancer drugs. As I describe in the remainder of this thesis, this screen harnessed nearly 400,000 compounds across the chemical libraries of the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH)[36]. We identified novel inhibitors with diverse structural scaffolds and single-digit μM potency in our bGUS biochemical assay.

CHAPTER II

VIRTUAL HIGH-THROUGHPUT SCREENING FOR DISCOVERY OF MICROBIOME ENZYME INHIBITORS TO ALLEVIATE CANCER DRUG TOXICITY

Summary

I present a holistic, first-in-kind application of vHTS of the 399,423 compounds within the NCGC chemical libraries towards discovery of a bGUS-inhibitory prodrug to improve the tolerability of standard-of-care (SOC) anti-cancer drugs among neoplastic patients. Herein, our virtual screening platform—with validation via a biochemical bGUS assay *in vitro*—revealed sixty-nine (69) compounds that associate to common scaffolds for the potential development of this adjuvant, all of which have significantly inhibitory activity for bGUS and do not cause eubacterial death. Potency analysis of our approach revealed a hit rate that is enriched 23-fold for a comparative non-virtual HTS task and confirmed a top-ranking hit with IC_{50} = 3.8 μM that may suggest an early signal for drug repurposing. Therefore, our results are highly translational and could inform the successful downstream development of a supplement to SOC chemotherapy that targets the human microbiome to prevent SAEs associated with first-line cancer treatments.

Methods

1. Identification of a representative bGUS analog

To develop a vHTS platform for discovering inhibitors of bGUS, we first identified an ortholog of bGUS against which we could screen the NCATS chemical library. Herein, we focused our vHTS efforts on "H11G11-BG," a representative bGUS ortholog implicated in the processing of regorafenib, a kinase inhibitor used to treat metastatic colorectal cancer,

hepatocellular carcinomas, and gastrointestinal stromal tumors[17,18,25]. The H11G11-BG protein was first identified as a bGUS gene encoded by the human fecal metagenome[17].

We isolated the *H11G11-BG* genetic sequence (NCBI Accession: CBJ55484) and purified the recombinant protein for biochemical hit validation, per the procedure of Wallace *et al*[22]. Briefly, we optimized the nucleotide sequence for translation in *E. coli* and inserted the gene into a pLIC-His vector with ampicillin resistance and linkage to an *N*-terminal 6x-Histidine affinity tag for downstream purification. We then transformed competent cells (*E. coli* BL21-DE3), grew them to an OD_{600} of 0.6, and induced gene expression with IPTG. Then, we centrifuged the cells and resuspended them in buffer. We poured cell lysate into a column pre-loaded with 5 mL His-NTA resin and collected the resulting fractions. We employed size exclusion as a final purification step and eluted protein isolate into 20 mM HEPES, pH 7.4, and 50 mM NaCl for biochemical activity assays. Then, we assessed purity by SDS-PAGE gel separation and gel filtration elution (>90% purity); we determined protein concentration using a spectrophotometer at a 280 nm absorption setting, using a standard combination of molecular weight, extinction coefficient, and Beer-Lambert Law calibration data to calculate concentration. The 754-residue amino acid (AA) sequence of H11G11-BG is cataloged in Universal Protein resource (UniProt) (https://www.uniprot.org/)[37] as D5GU71[38].

2. Development of a single-template homology model for bGUS

Having isolated a representative sequence of bGUS in FASTA format, I developed homology models to predict representative quaternary structures of this target before executing vHTS against it. To perform this modeling, I used the Molecular Operating Environment (MOE) (https://www.chemcomp.com/Products.htm)[39], a suite of industry-grade chemical computing

software for computer-aided molecular design and visualization of target structures, and imported the AA sequence of the protein into the Sequence Editor (SE) of MOE. To guide the selection of a template for homology modeling, I queried the H11G11-BG FASTA sequence in the Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi)[40]; this search gave the closest relatives of bGUS to be H11G11-BG homologs and homologs of β-galactosidase. At an arbitrary benchmark of 80% identity, there were nine (9) potential templates. I used this list of potential guides from BLAST to prioritize a shortlist of templates from suggested sequence alignments in MOE, adding the top five (5) hits of lowest *p*-value from two-sample similarity testing, as provided within the alignment window of the MOE user interface (UI) and that compared an input sequence with all available AA sequences from the most recent versions of the Protein Data Bank (PDB) (https://www.rcsb.org/)[41] and UniProt (as were available when I completed this study in Summer 2019). The five (5) hits I curated from MOE were also β-glucuronidases and β-galactosidases, in accordance with the BLAST results that I observed. Therefore, after selecting the top five (5) hits from MOE, along with those with overlap with the BLAST results and available from the UI, I aligned these prospective templates along the reference frame of our primary sequence within the MOE SE to allow for a more granular view of template similarity at each AA residue along the H11G11-BG chain. I then generated a point accepted mutation (PAM) matrix to identify the template chain with highest residue similarity to our sequence of H11G11-BG; this information facilitated my decision of a template for generation of a homology model, which I selected as a function of low *p*-value from the MOE UI, highest % identity and residue similarity from the PAM matrix, and homology between template length and the length of our bGUS sequence to allow for prospective coverage of the span of our target structure. Using these heuristics, I selected 6MVH (a flavin

mononucleotide-binding β-glucuronidase isolated from *Roseburia hominis*; https://www.rcsb.org/structure/6MVH)[42] as the most optimal template for bGUS; 6MVH presented 58.5% sequence identify to our query sequence, which is nearly double the benchmark 30% necessary for the confident generation of a predictive quaternary structure[43]. Therefore, Igenerated ten (10) putative homology models using 6MVH as a guide and the implementation of the Assisted Model Building with Energy Refinement force field that is inbuilt within MOE to provide property methods for the thermodynamic calculations that underlie standard, one-template homology modeling[39,44].

Upon studying the homology model of the highest rank from MOE, I identified the putative binding site in this model, per the co-localization of key active site residues for bGUS, harmonically identified by Wallace *et al*. as Tyr433, Tyr437, Phe448, Glu464, Lys563, Arg564[22]. Indeed, the observation of these residues or those of similar biochemical profile at these positions validated the strength of this homology model. Also supportive of the consistency of my approach with existing knowledge on bGUS was my observation that—upon alignment—the primary sequences of 6MVH and my homology model are highly similar for the 441 residues of closest adjacency to the template active site. However, for the remaining 313 residues of 6MVH, the homology model is strikingly dominated by extraneous alpha helices and turns that are not in keeping with the selected template. While these structures likely account for empirical reductions in similarity between my template and my homology model, I noted that they appear insignificant in the scope of a downstream vHTS task, as this mode of compound screening considers rigid docking of hit candidates only to a specified ligand binding site. Therefore, though this rationale may appear to neglect the dynamism and the structural importance of non-active site domains within our target, I assert that the high similarity of 441 residues near the

bGUS active site could—within the framework allowed by vHTS, which requires specification of a granular region of a target structure against which to screen compounds[34]—provide us hints at the dynamism that occurs upon ligand binding at the site where this flexibility may most significantly affect efficacy of competitive inhibition. However, I note—as is standard across most of the vHTS literature[31,32,34]—that the drug discovery task I report in this thesis involves thoroughly rigid docking. Finally, to confirm that this homology model was of sufficiently high quality to warrant vHTS, I generated a Ramachandran plot to note potential steric strain within the target structure. From MOE's Ramachandran plot, I found that my model produced an acceptably low number of residues (12) with outlying levels of angle strain. Then, I annotated the outliers on the ribbon structure of my homology model to observe the location of these residues. In doing this, I noted that most of these residues fell within a variable section of the target (i.e., in a loop or other functionally dynamic domain), suggesting that the few loci with excessive strain in my structure are of little consequence to the robustness of my model, as these regions may adopt allowable conformations to mitigate the strain in a true, functional *in vivo* system. Nonetheless, I note that that action site of my homology model—as well as its bGUS template—appears to be surrounded by several loops. While this suggests that the active site may be dynamic, I again affirm that our vHTS investigation is not compromised by this structural motif, as vHTS employs rigid docking.

3. Development of a multiple-template homology model for bGUS

The strength of a homology model is of paramount importance in developing a compelling vHTS platform—especially for tasks of precision drug design, such as that which I report in this thesis. Therefore, while I trusted the strength of the homology model I generated

from MOE, I also wished to develop the most biologically-representative model as I could. I note that MOE can only generate homology models from a single template; it cannot harness the power of newer *ab initio* methods, which can achieve harmonization of domain-specific modeling across several templates of high domain identity within a target AA sequence[45–47]. By selecting a combination of highly-representative guide structures, this "threading" approach has the potential to generate more accurate homology models than may be available through MOE, as it employs an ensemble of high-ranking structure guides, rather than one (1) holistic template that may not represent all domains of a target structure well. Indeed, for our first-pass attempt to discover competitive inhibitors of bGUS by vHTS, this rationale is especially relevant for our target's active site.

Therefore, recognizing the benefits that an *ab initio* approach to predicting the structure of bGUS could hold, I generated a second homology model of this target via "threading." While a discussion of the mechanistic details of this method is beyond the scope of this thesis, I used an open-source, "out-of-box" server to generate an *ab initio* homology model of bGUS, treating the quantum methods to generate this model as a "black box." I therefore assigned the AA sequence of H11G11-BG to the Iterative Threading ASSEmbly Refinement (I-TASSER) server, hosted by the Yang Zhang laboratory at University of Michigan (https://zhanglab.ccmb.med.umich.edu/I-TASSER/)[48] to generate a "threaded" model through similar querying and alignment of the target sequence with available structures from PDB and UniProt as I describe above. Following receipt of this homology model—as generated from six (6) domain-specific glucuronidase and glycosidase templates—I loaded the model into MOE for visualization. I noted that the strength of this model was apparent from the high degree of conservation between the active site residues that the I-TASSER server identified as a putative binding site within its self-generated homology

model and the active site residues as reported by Wallace and colleagues. Therefore, taking the domain of the H11G11-BG model that I-TASSER reported as its active site, I noted the following key residues: Trp510, Glu464, Tyr433, Lys531, Asn529, Asp132, His315, Glu376, Asp375. Further reflecting on the results returned by the I-TASSER server, I noted that this putative active site falls within the domain of a bGUS analog from *Rumnicoccus gnavus* (6MVG; https://www.rcsb.org/structure/6MVG)[49]. Additionally, I-TASSER provided us with quantitative metrics that cemented the similarity of my homology model and this top-ranking template, such that I could confidently extrapolate the template's active site to be homologous with that of my model: I-TASSER's confidence score (C-score) for the model's active site was 0.02, per a scale of C-score ϵ [-2, 5] (5 is most confident[50]), and the template modeling score (TM) was 72 ± 11%, per a range of TM ϵ [0, 100]% (100% is a perfect match between a model and an ensemble of guide structures).

Visualizations of MOE and I-TASSER homology models that I generated for H11G11-BG are available within Figure 2, below. Structure files for these models—along with our primary sequence of H11G11-BG, the template search and threading results from the I-TASSER server, and a Ramachandran plot for my MOE model—are available in the "Appendix" section of this thesis.

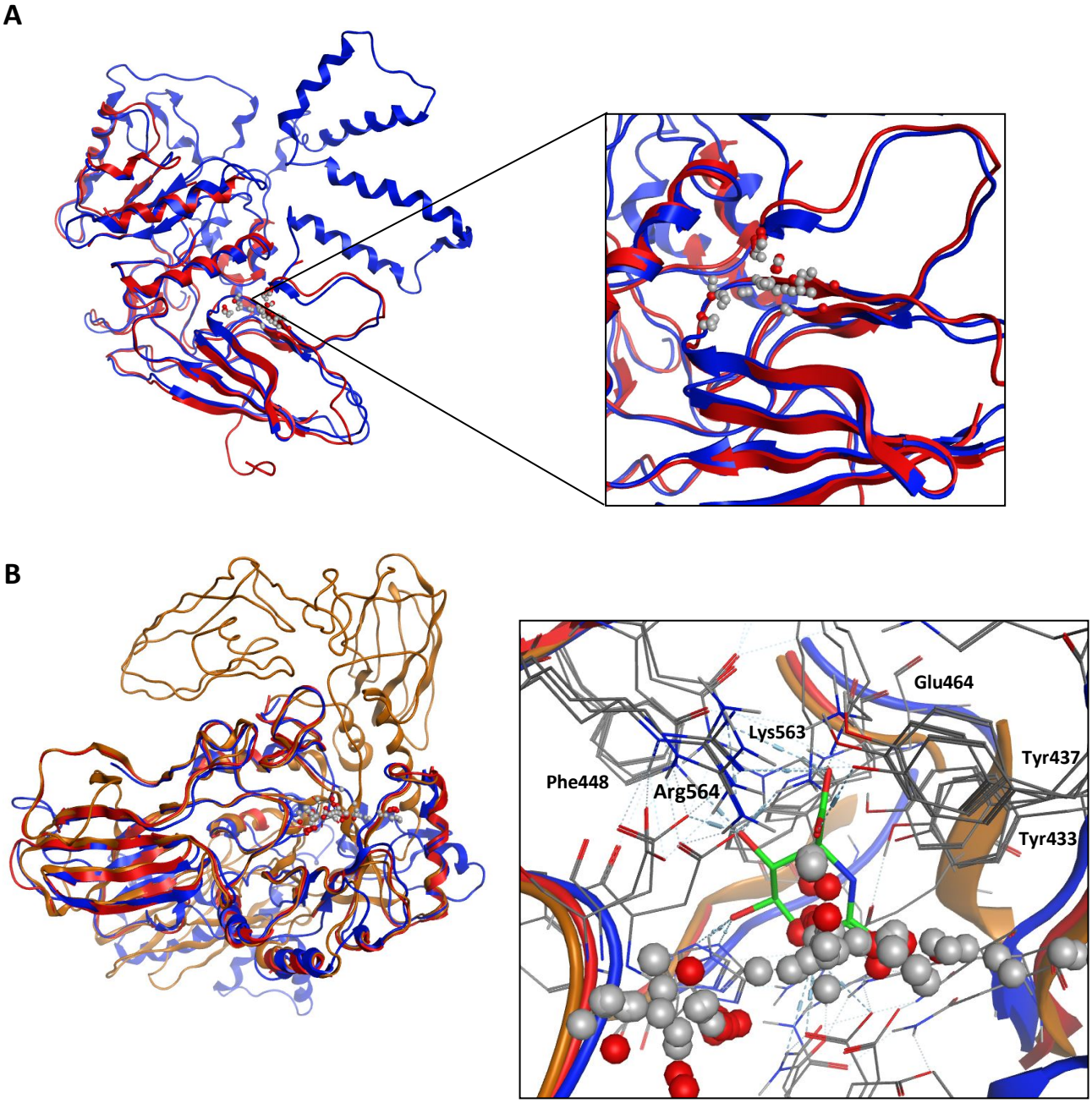

Figure 2: I developed homology models for bGUS using both MOE and the I-TASSER server. Panel A shows the single-template homology model I generated in MOE with an enlarged view of the model's predicted active site (red: MOE model, blue: 6MVH template). Panel B shows the threaded model I obtained from I-TASSER, superimposed on the homology model I generated in MOE and with an enlarged and residue-labeled view of the models' predicted active sites when occupied by an analog of glucose (gold: I-TASSER model, red: MOE model, blue: 6MVH template).

4. Docking analysis

Having generated robust homology models for bGUS and identified their putative active sites, I next sought to analyze the binding modes of common bGUS ligands available from the literature, both to understand key ligand binding interactions—whose presence I could use to gauge the success of a downstream vHTS platform—and to further confirm that the binding modes reported in the literature hold for the homology models I generated, as a further attempt at quality assurance for my target structures. In doing this, I first attempted to bind a known bGUS substrate, following by binding of known active site ligands, to discover and elucidate potential modulatory binding modes within my target model.

From PDB, I first downloaded the glucaro-*d*-lactam ligand (3K4D; https://www.rcsb.org/structure/3K4D)[51] and uronic isofagamine (5Z1A; https://www.rcsb.org/structure/5Z1A)[52] inhibitor to ensure reproducibility of these binding poses, as given in the PDB template entries via interaction of the ligand with the key residues outlined above. Noting the maintenance of the associated key residues (per Wallace and colleagues)—including hydrogen bonding with residues like Lys563 and Tyr433, as well as proximity for nucleophilic attack at Glu464—I then queried PDB by the term "glucuronidase" and downloaded all nine (9) unique ligands that resulted from this search, noting similar results as above.

Armed with the maintenance of these interactions, I prepared for virtual screening. Since both homology models appeared to provide a high degree of fidelity—and because I sought to perform screening against ~400,000 compounds—I decided to move forward with the I-TASSER model for our drug discovery analysis. Given the inherent strength of this model's *ab initio* design—as well as the satisfactory degree of homology I established between the model

and its highest-ranking bGUS template—I decided to optimize the efficiency of our vHTS approach by screening solely against my I-TASSER structure, instead of adopting the computationally-intensive task of screening against both homology models in parallel.

It is important to note that vHTS requires coordinated learning between both a user and his/her machine. The user must specify a binding mode that screening candidates should replicate in order for them to rank highly on a list of hits, and he/she must also create a "gridbox" containing the target residues that maintain the key interactions that ligands interacting in this binding mode should maintain. Therefore, before a user begins virtual drug screening, it is essential for him/her to identify the specific form of the target structure against which he/she wishes to screen and to select a docked ligand that has an appropriate binding mode to guide the identification of vHTS hits. In this regard, it is also of paramount importance to consider the differences between vHTS on *apo* and *holo* target structures if there is evidence of cooperative binding within the previous literature about the target, as the bound and unbound configurations of the target are likely to vary non-trivially.

Therefore, as a preliminary attempt at preparing our target structure for drug screening, I considered the most optimal binding mode for vHTS to be that which maintains the key interactions I noted from the literature, while also demonstrating a depth of interaction with the bGUS active site that appears to cover a maximum of its surface area. Furthermore, while the above steps centered around the development of an *apo* model, I also undertook a systematic search of the bGUS literature to determine the relevance of vHTS against a *holo* representation of our target structure.

Following a review of available knowledge on cooperative bGUS MOA, I identified literature demonstrating that the bGUS strain from *Bifidobacterium dentium* (*Bifd*) is known to

bind competitive inhibitors cooperatively and in a substrate-dependent manner, with inhibitors binding at a site immediately adjacent to the active site while maintaining key stabilizing and reactive interactions with the active site residues I identified[18]. This suggested to us that I should specify a *holo* binding mode to generate accurate vHTS results. Therefore, I prepared my bGUS homology model using both a sugar substrate and a large ligand capable of low-energy binding at a site immediately adjacent to the active site of bGUS.

To generate this representation, I downloaded a β-(*D*)-glucuronic acid substrate molecule from the library of available ligands in PDB. Then, in MOE, I docked this molecule to my I-TASSER model to generate the appropriate target structure against which I could screen candidate inhibitors in a substrate-dependent manner. Next, to generate an optimal binding pose for application in virtual screening, I curated a library of potential inhibitors, so I could select a template from among these compounds. To do this, I first reflected on previous *in vitro* assays that our group at NCATS had attempted to discover a battery of bGUS inhibitors. I downloaded the ~35 SDF files of all compounds that we previously screened against other non-H11G11-BG bGUS strains and which demonstrated inhibitory activity ($IC_{50} < 20$ μM, efficacy < -50%, curve class (CC) ϵ [-1.1, -1.2, -1.3, -1.4]) and loaded them as a library into MOE. I then harmonized this set of compounds with the ligands I queried from PDB and enabled MOE to dock these compounds to our *holo* target structure. Then, to prioritize these ligands for generation of the most optimal binding pose, I ranked the compounds by compatibility with the *holo* target, ordering the hits by ascending, MOE-generated docking score (a proxy for the predicted energetic change of the target system upon docking of each compound).

Upon prioritization of my docking hits, I observed that the compound with the lowest MOE docking score was a sulfonated glucose molecule; however, given that this compound

closely mimics bGUS substrate—and presents with small surface area around the bGUS active site—I decided to probe compounds with larger surface area and slightly higher binding energy differentials in selecting a template binding mode. Doing this, I determined that the lowest binding energy ligand that also met my requirements for coverage of key residues was 1-((6,8-dimethyl-2-oxo-1,2-dihydroquinolin-3-yl)methyl)-1-(2-hydroxyethyl)-3-(4-hydroxyphenyl)thiourea, a ligand I extracted from PDB and that corresponds to a known inhibitor of *Escherichia coli* bGUS (5CZK; https://www.rcsb.org/structure/5czk)[53]. Re-docking this ligand to our *holo* target structure and using MOE's *Protonate3D* functionality to obtain the least sterically-encumbered conformation for all target residues with rotatable bonds, I captured the resulting binding pose as my template for virtual screening. I noted that this pose appears to maintain the key interactions for our bGUS target. For my readers' review, I include a visualization of this pose as Figure 3, below. I generated Figure 3 via the open-source structure visualization software Chimera, which is hosted by the University of California, San Francisco (https://www.cgl.ucsf.edu/chimera/)[54].

Generally, it is important to note that this docking exercise was not deterministic and that other binding poses may have also been appropriate templates for vHTS. Therefore, my intent in applying the heuristics I employed and describe above was to discover one of these appropriate binding poses through strategic application of foundational biochemistry and computational design. To aid comprehension of the target modeling methods I describe above, I provide a graphic summary of my approach in Figure 4, below.

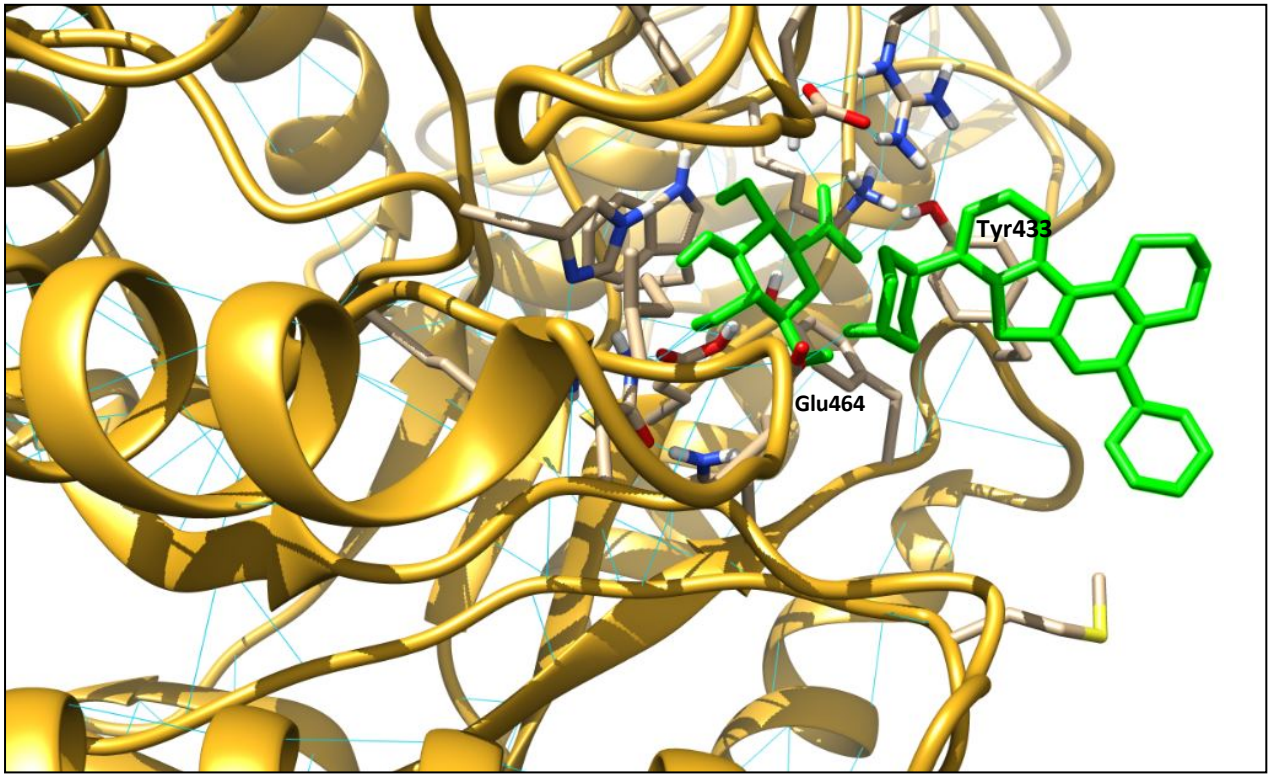

Figure 3: From my docking experiments, I developed a *holo* template of bGUS bound to 1-((6,8-dimethyl-2-oxo-1,2-dihydroquinolin-3-yl)methyl)-1-(2-hydroxyethyl)-3-(4-hydroxyphenyl)thiourea for development of my gridbox and subsequent vHTS. This template inhibitor appears to exhibit mixed-mode inhibition, as its most favorable binding pose occupies a site tangent to the bGUS active site but maintains several of the key interactions that I would expect to observe under competitive inhibition. Constraining residues that I specified in my gridbox are labeled, and both the substrate glucose analog and the template inhibitor—with its solvent side exposed—are colored green.

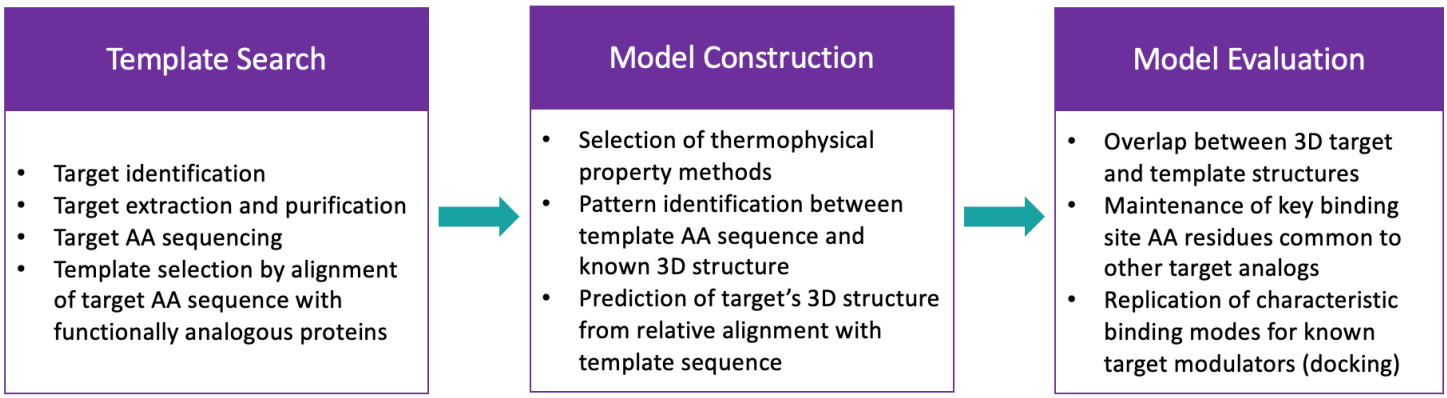

Figure 4: A summary of the homology modeling approach I developed to predict a 3D structure for our H11G11-BG analog of bGUS, which considers both structure prognostication and validation via docking experiments

5. vHTS

Next, I prepared to screen the 399,423 compounds within the NCGC chemical libraries against the *holo* representation of my homology model that I prepared via my docking experiments. Compound prioritization in a vHTS platform hinges on the replication of a user-specified binding mode in establishing a definition of energetic favorability; therefore, to prepare for vHTS against bGUS, I defined the active site-adjacent ligand binding region I observed as a guide for my screens, ensuring to include the key residue interactions with my template inhibitor as *constraints* for the calculation of binding energy and the prioritization of screening hits. To prepare our target structure accordingly, I employed version 3.3.1 of the *OEDocking* package, which is available through OpenEye Scientific Software (https://www.eyesopen.com/oedocking)[55].

Therefore, using the *MakeReceptor* method within this package, I defined a gridbox to include the key residues within the configuration of the I-TASSER homology model that resulted from binding of substrate to the bGUS active site; I removed the substrate molecule itself before screening to avoid steric penalties from artifactual binding patterns that could result in collisions between the ligand and substrate molecules (OEDocking tests several hundreds of binding poses in the virtual screening of each candidate in a library against a target). Indeed, the first stage in vHTS is to discover the most favorable binding position for each screening candidate; then, the most energetically-favorable hits may be prioritized by binding energy. The key residues—and their associated interactions—that I included in the gridbox were the following and were in accordance with existing literature on bGUS MOA that I discussed in earlier sections of this thesis; residues that I specified as constraints are **bolded**.

- **Glu464** (This residue allows H-bonding & nucleophilic attack, depending on the ambient pH of the target-ligand ensemble.)
- **Tyr433** (Given the relative positioning of phenyl rings within this residue and within my template inhibitor, meta-stabilization via π-π stacking is possible if rotation of the Tyr R group places its phenyl ring parallel to the candidate's phenyl ring. Indeed, I observed this positioning in my MOE and MakeReceptor visualizations.)
- Tyr437 (This residue allows for stabilization of the target-ligand assembly via H-bonding to the residue's free hydroxy group.)
- Lys563 (This residue allows for stabilization of the target-ligand assembly via H-bonding to its charged R group.)
- Arg564 (This residue allows for stabilization of the target-ligand assembly via H-bonding to its charged R group.)
- Phe448 (This residue appears to provide steric hinderance to inhibitor binding in the *apo* configuration of bGUS, but substrate binding facilitates its retrograde movement, such that this configurational change allows ligands to access the inhibitor binding site that lies adjacent to the target's active site.)

Then, from the command line, I harmonized the 399,423 compounds within the NCGC chemical libraries with my docking collection of ligands I extracted from PDB and the most active compounds from our previous bGUS assays. I intentionally designed this list to contain redundancy (as our self-selected docking library is an inherent subset of the NCGC libraries) to facilitate a "sanity check" on my vHTS results: if vHTS executes correctly, redundant cherrypicks and PDB ligands should be listed repeatedly and at the top of the list of hits.

Next, I executed vHTS from the command line, using the *FRED* method within *OEDocking*. This protocol allows for the generation of several hundreds of conformers of each screening candidate and fast and rigid docking of these conformers to the target gridbox that I previously defined. Therefore, given the ostensibly high dimensionality and time complexity of this screening task, I ran vHTS within a high-performance computing cluster on Biowulf, a NIH server capable of supercomputing[56]. To optimize the efficiency of candidate screening via FRED, I converted all drug structure files to the file format *OEB.gz* before executing vHTS on Biowulf.

I present our vHTS outcomes in the succeeding "Results and Discussion" section of this thesis. In this section, I also include summary graphics of the above methods that present key numbers from—and the relative positionioning of—these workflows within the overall framework of our investigation.

6. Validation of vHTS results in vitro

To confirm the activity of our 291 cherrypicks, we performed a biochemical validation assay and a subsequent counterscreen, using a strain of bGUS that is homologous to H11G11-BG. This test employed the standard 4-methylumbelliferyl-β-D-glucuronide hydrate (4MUG) bGUS inhibitory assay, which allows for determination of candidate inhibitors' IC_{50} values by spectroscoptic titration, during which the intensity of fluorescent 4MUG substrate under different inhibitor dose conditions permits quantification of unconverted substrate concentration (a proxy for enzyme functionality) as a function of inhibitor concentration[57,58]. Before testing our cherrypicks, we purchased 4MUG (type # M9130[59]) from Sigma-Aldrich, Inc. (St. Louis, MO, USA) and validated the above assay in-house with a collection of 174,602 compounds within our

chemical libraries, screening all compounds through robot-assisted HTS in a five-point dose response series at concentrations between 0.5 μM and 114.0 μM and working conditions of 0.5 nM bGUS concentration and 100.0 μM 4MUG per well. Before inhibitor testing, we allowed both enzyme and substrate to incubate for 1 hour in a standard mixture of 50 mM HEPES buffer (pH 7.5) and 0.01% Triton X-100 detergent solution. During each titration, we used a ViewLux spectrophotometer[60] to obtain 4MUG intensity data from each of our 1,536-well plates. Then, after confirming high signal-to-background ratio (S:B) and Z'-factor for this workflow, we replicated the assay on our cherrypicks, several of which we did not originally screen through our rapid assay validation procedure.

In the following "Results and Discussion" section of this thesis, we discuss the results of our validation assay for our top vHTS hits and dissect the significance of these outcomes when taken in complement with the informatics workflow we present above.

Results and Discussion

1. Results from virtual screening

Biowulf required three (3) days to complete the vHTS job I describe in "Methods," which received top priority in the queue for server access. Upon receiving 2,500 low-binding energy hits from Biowulf, I noted replication of PDB and HTS client cherrypicks, as I expected, and therefore proceeded in applying a series of biochemical heuristics to prioritize the top hits for further interrogation. First, armed with the list of hits ranked ascendingly by binding energy, I arbitrarily selected the top 1,000 compounds for manual curation (a sample set of these compounds is available within the Supplement). Then, as a mechanism of "quality control," I removed all compounds resembling broad-spectrum antibiotics, such as cephalosporins,

fluoroquinolones, and tetracyclines, given our interest in discovering precision bGUS inhibitors and the relevance of this goal to the immunological manifestations of neoplasia, as I describe in the "Introduction" section of this thesis. Additionally, I removed lower-ranking compounds with bulky moieties, given my concerns about potential steric hinderance at the inhibitor binding site, as I discovered through docking experiments on our *holo* target model. I also parsed compounds with problematic chemistry, such as hits with several cyclopropane functionalities (given their associated angle strain and instability) and hits with trifluorinated aromatic moieties (given the strong electron-withdrawing effects of these groups and their potential impacts on efficacy). I also eliminated other hits with functionalities that are strongly associated with cytotoxicity on an *ad hoc* basis and per the expertise of our study team.

Next, to arrive at a list of high-priority hits for validation within a relevant biochemical assay, we circulated the list among the computational chemists and biologists within our study team to select chemical series within this list that appeared promising for further interrogation of efficacy and potency; I harmonized their selections, resulting in 291 cherrypicked compounds with low binding energy for our *holo* target model. I then compiled a library of structure files for these compounds and imputed them to MOE for re-docking to my homology model, to ensure that these cherrypicks maintained the key interactions I specified in the creation of my vHTS gridbox. In doing this, I noted that our prioritized hits interacted with the key stabilizing and nucleophilic residues I identified within the bGUS active site and demonstrated extensive surface area coverage of this domain, as several top-ranking compounds contained multiple six-membered rings. This motif promoted jutting of the ligand towards the solvent side of the target, as appears common with many of the effectual ligands for homologous strains of GUS that I found in PDB's ligand library.

By visual inspection, I also noticed the recurrence of three (3) distinct motifs, at large, within our high-ranking cherrypicks: a pyrazole group, a benzoimidazole derivative, and multiple electron-donating halogens substituted and distributed across component benzyl rings of these ligands.

Nonetheless, I did not observe the presence of a π-π stacking interaction with Tyr433 across the optimal binding poses of our top-scoring candidate inhibitors. This might be the outcome of flexible, rotatable bonds at the residue-aligned sites within the ligand structures; alternatively, this phenomenon could have resulted from an inbuilt limitation in the sensitivity of detecting a π-π stacking within OpenEye's constraint recognition system, as the presence of several atoms in appropriate proximity and approximately parallel to Tyr433 might have triggered faulty recognition of this interaction in the generation of our list of hits. Nonetheless, I maintained our list of 291 top-ranking cherrypicks without prejudice, to evaluate their comparative performance in biochemical validation assays *in vitro*.

2. *Results from validation assays* in vitro

We observed strong performance for the bGUS 4MUG assay that we validated in high-throughput form, as we obtained S:B = 123 ± 9 and Z'-factor = 0.93 ± 0.04 (Z' ϵ [0,1]; Z'=1 gives the highest possible HTS assay performance[61]) from our in-house testing. In analyzing the results of this assay and its counterscreen for our vHTS cherrypicks, we identified potency for sixty-nine (69) of our 291 trial compounds (i.e., CC ≠ 4 and CC < 0), giving a 24% hit rate for our drug discovery exercise. Among these potent compounds, we observed few singleton hits. Additionally, we uncovered thirteen (13) potent compounds with activity heuristically defined as "industry-grade" for future drug development (i.e., IC_{50} < 10 μM, efficacy < -50%)[62].

3. *Discussion of key results*

From our coupling of vHTS and biochemical assays to predict bGUSi efficacy, I uncovered sixty-nine (69) potent and efficacious compounds with inhibitory action on bGUS, predicting a new structure of a bGUS homolog resident to the human gut microbiome and then screening a large compound library *in silico* against this homology model to identify series with logical chemistry and the highest alignment with the target's rationale. We then validated our drug efficacy hypotheses in a representative system *in vitro*. Thirteen (13) of our hits map to scaffolds that demonstrate "industry grade" bioactivity. A visual summary of this protocol—as well as the numerical results arising from its execution—are available in Figure 5, below.

I believe that the results I present here represent a major stride in evaluation of the druggability of a target implicated in adverse reactions to high-use, first-line anti-cancer drugs, as we obtained a 24% hit rate resulting from our drug discovery experiments. Given that most assays performed under HTS conditions result in hit rates ~1%, our much higher rate suggests that our approach to discovery of a bGUS inhibitor presents with nearly 23-fold enrichment over a similar method of non-virtual drug discovery.

It is important to note that manual evidence synthesis is an important component of virtual drug discovery. While the "automated" components of our drug discovery process provided the edifice upon which I could specify the minimum biochemical specifications to define a "hit" (and thereby eliminate the majority of our large screening library from consideration), the procedure for ensuring practicality and efficiency of our platform (i.e., our prioritization of the most relevant hits for downstream testing) relied on our ability to curate expert knowledge. vHTS is a hypothesis-generation tool that allowed us several machine-defined actives to consider manually for signals of efficacy. However, the strength of our platform—as

evidenced by its high hit rate and inbuilt bGUS specificity—is holistically associated to our power to identify the most promising hypotheses of drug activity from our vHTS results and our biochemical reasoning. This information allowed for us to observe activity for a high percentage of our prioritized predictions within our validation assay. Therefore, I believe that our hit rate is neither artifact nor artificially-inflated. Instead, it speaks to our selection of powerful heuristics for prioritizing vHTS hypotheses.

I similarly affirm that the best application of evidence synthesis to vHTS results is neither arbitrary nor deterministic, as the use of strong conceptual logic and *ad hoc* decision-making in evaluation of each listing on an automated bioactives list speaks to a design-oriented approach to drug discovery.

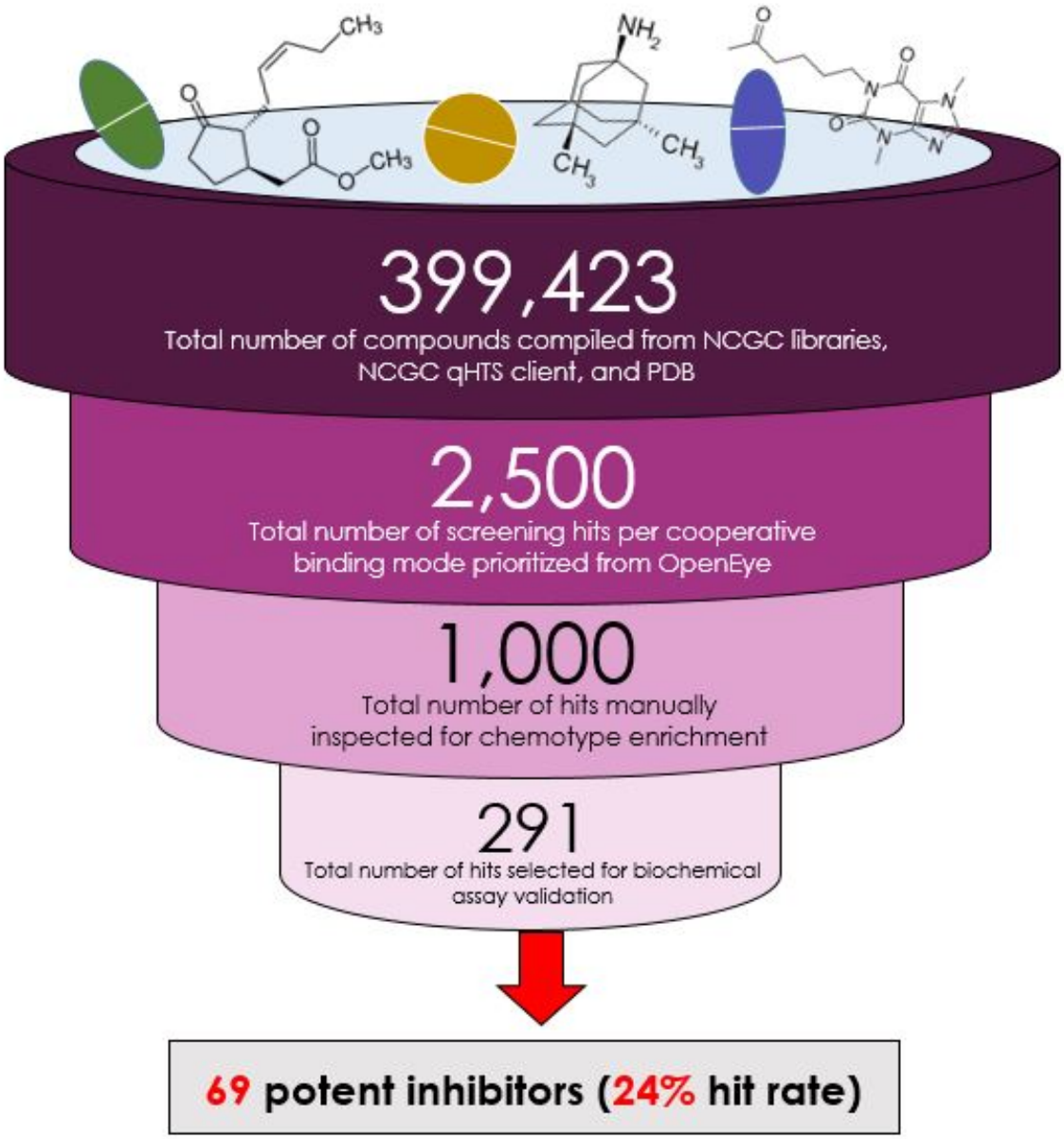

Figure 5: A visual summary of the workflow we employed to discover bGUS inhibitors via vHTS and a validative biochemical assay, accompanied by quantification of this platform that supports our claims of systematic and high-throughput drug screening, sensitive heuristics for prioritization of vHTS hypotheses, and a high hit rate

Within our list of sixty-nine (69) hits with confirmed activity *in vitro*, I further prioritized this library by observations of "industry-grade" bioactivity (IC_{50} < 10 μM, efficacy < -50%); in Figure 6, below, I present the structure, binding mode, and dose-response behavior of the top three (3) hits from this shortlist. We generated dose reponse curves using version 8.2.0 of

Prism[63]. I note adherence of the binding modes among these compounds to my core criteria in selecting inhibitor templates through my docking experiments, which served as preliminary proof-of-concept of the efficacy of our approach and a further "sanity check" in confirming the validity of our results.

Because I assert that our sixty-nine (69) top hits cluster among representative chemical series (and therefore present few singletons), here, I discuss the most potent compounds within each of three (3) highest-ranking series.

As Figure 6 shows, NCGC00253873—our top hit and a representative compound of the most potent chemical series within our results—shows maintenance of key interactions with the "constraining" residues that I specified in my vHTS gridbox when I re-docked it to my I-TASSER homology model. It demonstrates IC_{50} = 3.8 μM for the physiologically-relevant strain of bGUS that we employed in our biochemical assay, with efficacy = -71%, CC = -1.2, and a clearly-defined, logistic dose-response curve. This compound is an analog of the RUC-2 inhibitor—a variant of the series of RUC-4 inhibitors (RUC-4i), which have garnered recent interest for their anti-thrombogenic inhibition of glycoprotein IIb/IIIa[64–66]. To date, RUC-4i has completed a phase I randomized, controlled trial of dose tolerability (https://clinicaltrials.gov/ct2/show/NCT03844191?term=ruc-4&draw=2&rank=2)[67] and is now in the beginning stages of a phase II open label trial for pharmacodynamic and pharmacokinetic characterization of efficacy in reducing disease severity among patients with sinus tachycardia-elevation myocardial infarction (https://clinicaltrials.gov/ct2/show/NCT04284995?term=ruc-4&draw=2&rank=1)[68]. Therefore, this result might provide an early signal for a drug repurposing opportunity within the therapeutic space I describe in this thesis. Additionally, our

finding may be especially timely for further evaluation, given the present interest in treatments of coagulopathy[69–71] like the RUC-4i series.

A representative of our second highest-ranking series—5-(2,3-dichlorophenyl)-*N*-(3-(oxazolo[4,5-*b*]pyridine-2-yl)phenyl)furan-2-carboxamide (NCGC00099510)—showed IC_{50} = 9.6 μM for bGUS, with logistic dose-response behavior, CC = -1.2, and efficacy = -70%. A third series-representative hit compound I present in Figure 6—4-(cyclopropylmethyl)-*N*-phenyl-2,3,4,5-tetrahydro-1*H*-pyrido[4,3-*e*][1,4]diazepin-8-amine (NCGC00411059)—also demonstrated logistic dose-response behavior, with IC_{50} = 11.1 μM, efficacy = -72%, and CC = -2.2. The inclusion of this compound in our list of top-ranking compounds speaks to the power of vHTS in providing an additional layer of "quality control" to standard HTS workflows, as, in our former non-virtual, HTS probes of bGUS analogs similar to H11G11-BG, we screened this compound against bGUS in triplicate and with a five (5)-point titration series to develop a full dose-response curve, but we did not observe its bioactivity. Therefore, we had de-prioritized NCGC00411059 from future consideration as a potential bGUS inhibitor, but the strength of our vHTS results now suggest to us that we should re-examine this candidate as it may have more therapeutic potential than our previous assessments of its activity revealed. We regard the observation of compound efficacy in our biochemical bGUS assay as powerful, since the significantly reduced scale of this probe compared to that of the HTS platform at NCATS[72] suggests that its results may be viewed as more definitive and less "hypothesis-generating" than would otherwise arise from a standard HTS workflow. Furthermore, the situation I present of newfound observations of compound activity speaks to the power of vHTS in rescuing "false negatives" from a high-throughput drug discovery platform and therefore increasing its sensitivity. This is a pattern that is only scarcely observed in previous literature[73], but I believe it

to be an important attribute of the vHTS approach and therefore a notable aspect of the results i present.

I again note that nearly all of our screening hits demonstrated common, salient chemotypes, such that our sixty-nine (69) actives presented with few component singleton hits. The screening data from our bGUS biochemical assay for these compounds are available within the "Appendix" section of this thesis.

Finally, in importing our bioactive hits to MOE and re-docking them to my I-TASSER target model, I observed that these compounds all exhibited mixed-mode inhibition, as they bound to our bGUS homolog at a location adjacent to the active site while still maintaining interactions with key active site residues, as with the inhibitors I display in Figure 6. These observations foment the conclusions I describe above and present an interesting exploration of a new, substrate-dependent mode of bGUS inhibition that allowed us to discover several prospectively inhibitory series of precise therapeutic adjuncts. Therefore, the results I present in this thesis have potential application towards downstream development of a bGUS-inhibitory drug to accompany small molecule anti-cancer therapies and reduce the incidence of associated SAEs.

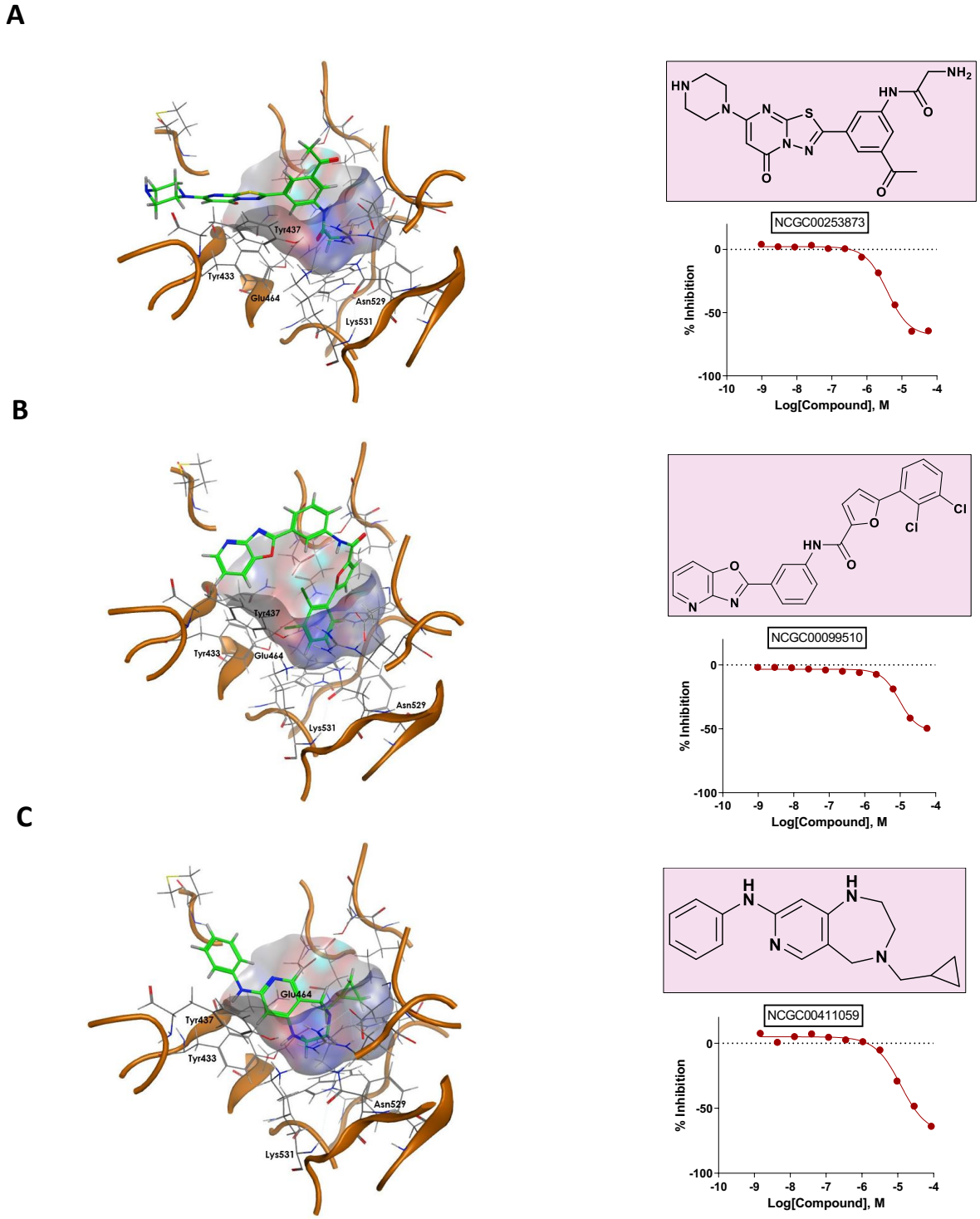

Figure 6: Representatives of chemical series among our top validated vHTS hits include RUC-2i (Panel A) and 5-(2,3-dichlorophenyl)-*N*-(3-(oxazolo[4,5-*b*]pyridine-2-yl)phenyl)furan-2-carboxamide (Panel B). Panel C provides an example of false negative rescue of 4-(cyclopropylmethyl)-*N*-phenyl-2,3,4,5-tetrahydro-1*H*-pyrido[4,3-*e*][1,4]diazepin-8-amine via vHTS. In each delineation of inhibitor binding mode, the substrate molecule is not visible (for ease of visualization), the ligand is colored green, and the binding pocket is shaded grey. Additionally, key active site residues are annotated; Phe448, Lys563, and Arg564 are in retrograde in the inhibitor binding modes I discovered.

Conclusions

SOC for patients with non-excisable neoplasia is prescription of an anti-cancer regimen; although most small-molecule anti-cancer drugs are well-established and demonstrate potency for their targets, patient non-adherence to these medications due to their associated SAEs remains a significant barrier to their efficacy. Gut commensal bGUS enzymes can reverse compound inactivation catalyzed by host phase II glucuronidation by cleaving drug-glucuronide conjugates and forming toxic levels of reactivated drug in the gastrointestinal tract. Thus, microbiota bGUS has been implicated as a novel drug target to prevent the gastrointestinal toxicities of existing therapeutics, including the anti-cancer drugs irinotecan and regorafenib[18,21,22,74–76]. However, as the gut microbiome provides immunological and digestive benefits to neoplastic patients who are often immunocompromised and present with iatrogenic epithelial cytopenia in their GI tracts, broad-spectrum antibiotics have a minimal therapeutic index for this population. This suggests that there is an unrequited need for a precision, bGUS-inhibitory adjunct therapy to be delivered alongside SOC anti-cancer drugs to ensure that neoplastic patients can better tolerate their treatment. Such a therapy could maximize the therapeutic value of existing anti-cancer agents, while simultaneously minimizing the safety risks of these drugs.

The work I present in this thesis presents a holistic, multilayered, and first-in-kind attempt at systematic discovery of hits for a bGUS inhibitor, which we attempt via the powerful approach of vHTS with downstream biochemical assay validation *in vitro*. The results that I present from analysis of ~400,000 compounds suggest success of this approach, as they center around a hit rate that is 23-fold enriched compared to standard HTS drug discovery initiatives, sixty-nine (69) hits that map to common scaffolds for bGUS inhibitor design (with thirteen (13)

hits demonstrating "industry-grade" bioactivity), and an opportunity to consider drug repurposing with a class of agents that is of increasing interest for its potential application in treating coagulopathy.

In the future, we hope to further develop our bGUS drug design platform by more detailed interrogation of druggable sites within our target model. While, in the work I present here, Iidentified candidates for inhibitor design through probes of candidate binding near the bGUS active site, we are curious to attempt our methods towards discovery of potential allosteric modulators of this target and, if an allosteric binding site appears plausible, to test the compounds we identified as hits from our current work to identify the potential for multimodal inhibition of this target. We also plan to further develop our testing platform of top candidates from our list of validated bioactives, as this will allow us to continue prioritizing our sixty-nine (69) hits. We are currently considering the development of biomarker discovery studies for bGUS inhibition and may consider testing our compounds in more representative *in vitro* systems, such as 3D human tissue culture, to better simulate their effects on the diverse microenvironments of human gut. We are also interested in developing appropriate mechanistic and enzyme kinetics probes to confirm our hypotheses on the modes of hit action that we developed through docking experiments and interpretation of the signals from our biochemical assay. Indeed, if the results that we present here hold true in these future screens, we believe that our platform could have strong potential in catalyzing the pace of drug discovery for a precision therapeutic that has application to improving health outcomes for millions of prospective patients[77].

REFERENCES

1. Pulley, J. M. *et al.* Using What We Already Have: Uncovering New Drug Repurposing Strategies in Existing Omics Data. *Annu. Rev. Pharmacol. Toxicol* **60**, 333-352 (2019) .
2. Challa, A. P. *et al.* Systematically Prioritizing Candidates in Genome-Based Drug Repurposing. *Assay Drug Dev Technol* **17**, 352–363 (2019).
3. Challa, A. P., Madu, C. O. & Lu, Y. BRCA 1/2 Tumors and Gene Expression Therapy for Breast Cancer Development and Metastasis. *Oncomedicine* **2**, 132–137 (2017).
4. Gospodarowicz, M. *et al.* Cancer Services and the Comprehensive Cancer Center. in *Cancer: Disease Control Priorities, Third Edition (Volume 3)* (eds. Gelband, H., Jha, P., Sankaranarayanan, R. & Horton, S.) (The International Bank for Reconstruction and Development / The World Bank, 2015).
5. Shewach, D. S. & Kuchta, R. D. Introduction to Cancer Chemotherapeutics. *Chem Rev* **109**, 2859–2861 (2009).
6. United States National Library of Medicine. *How does chemotherapy work? InformedHealth.org [Internet]* (Institute for Quality and Efficiency in Health Care (IQWiG), 2019).
7. Nurgali, K., Jagoe, R. T. & Abalo, R. Editorial: Adverse Effects of Cancer Chemotherapy: Anything New to Improve Tolerance and Reduce Sequelae? *Front Pharmacol* **9**, (2018).
8. Pearce, A. *et al.* Incidence and severity of self-reported chemotherapy side effects in routine care: A prospective cohort study. *PLoS One* **12**, (2017).
9. de Boer-Dennert, M. *et al.* Patient perceptions of the side-effects of chemotherapy: the influence of 5HT3 antagonists. *Br J Cancer* **76**, 1055–1061 (1997).

10. Geynisman, D.M. & Wickersham, K.E. Adherence to Targeted Oral Anticancer Medications. *Discov Med* **15**, 231–241 (2013).
11. daCosta DiBonaventura, M., Copher, R., Basurto, E., Faria, C. & Lorenzo, R. Patient Preferences and Treatment Adherence Among Women Diagnosed with Metastatic Breast Cancer. *Am Health Drug Benefits* **7**, 386–396 (2014).
12. McCue, D. A., Lohr, L. K. & Pick, A. M. Improving Adherence to Oral Cancer Therapy in Clinical Practice. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* **34**, 481–494 (2014).
13. Allain, E. P., Rouleau, M., Lévesque, E. & Guillemette, C. Emerging roles for UDP-glucuronosyltransferases in drug resistance and cancer progression. *British Journal of Cancer* **122**, 1277–1287 (2020).
14. Cummings, J. *et al.* Glucuronidation as a Mechanism of Intrinsic Drug Resistance in Human Colon Cancer: Reversal of Resistance by Food Additives. *Cancer Res* **63**, 8443–8450 (2003).
15. Hu, D. G., Mackenzie, P. I., Lu, L., Meech, R. & McKinnon, R. A. Induction of Human UDP-Glucuronosyltransferase 2B7 Gene Expression by Cytotoxic Anticancer Drugs in Liver Cancer HepG2 Cells. *Drug Metab Dispos* **43**, 660–668 (2015).
16. Pollet, R. M. *et al.* An Atlas of β-Glucuronidases in the Human Intestinal Microbiome. *Structure* **25**, 967-977.e5 (2017).
17. Gloux, K. *et al.* A metagenomic β-glucuronidase uncovers a core adaptive function of the human intestinal microbiome. *PNAS* **108**, 4539–4546 (2011).
18. Ervin, S. M. *et al.* Targeting Regorafenib-Induced Toxicity through Inhibition of Gut Microbial β-Glucuronidases. *ACS Chem. Biol.* **14**, 2737–2744 (2019).

19. United States Food and Drug Administration. Raloxifene package insert.
20. Chen, S. *et al.* Intestinal glucuronidation protects against chemotherapy-induced toxicity by irinotecan (CPT-11). *Proc Natl Acad Sci U S A* **110**, 19143–19148 (2013).
21. LoGuidice, A., Wallace, B. D., Bendel, L., Redinbo, M. R. & Boelsterli, U. A. Pharmacologic Targeting of Bacterial β-Glucuronidase Alleviates Nonsteroidal Anti-Inflammatory Drug-Induced Enteropathy in Mice. *J Pharmacol Exp Ther* **341**, 447–454 (2012).
22. Wallace, B. D. *et al.* Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. *Chem Biol* **22**, 1238–1249 (2015).
23. Awolade, P. *et al.* Therapeutic significance of β-glucuronidase activity and its inhibitors: A review. *European Journal of Medicinal Chemistry* **187**, 111921 (2020).
24. Biernat, K. A. *et al.* Structure, function, and inhibition of drug reactivating human gut microbial β-glucuronidases. *Sci Rep* **9**, 825 (2019).
25. Dashnyam, P. *et al.* β-Glucuronidases of opportunistic bacteria are the major contributors to xenobiotic-induced toxicity in the gut. *Sci Rep* **8**, (2018).
26. Pellock, S. J. *et al.* Three structurally and functionally distinct β-glucuronidases from the human gut microbe Bacteroides uniformis. *J Biol Chem* **293**, 18559–18573 (2018).
27. Bhalodi, A. A., van Engelen, T. S. R., Virk, H. S. & Wiersinga, W. J. Impact of antimicrobial therapy on the gut microbiome. *J Antimicrob Chemother* **74**, i6–i15 (2019).
28. Marchant, J. When antibiotics turn toxic. *Nature* **555**, 431–433 (2018).
29. Francino, M. P. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. *Front Microbiol* **6**, (2016).

30. Cully, M. Antibiotics alter the gut microbiome and host health. *Nature Research* (2019) doi:10.1038/d42859-019-00019-x.

31. Bajorath, J. Integration of virtual and high-throughput screening. *Nature Reviews Drug Discovery* **1**, 882–894 (2002).

32. Zoete, V., Grosdidier, A. & Michielin, O. Docking, virtual high throughput screening and in silico fragment-based drug design. *Journal of Cellular and Molecular Medicine* **13**, 238–248 (2009).

33. Hu, X. *et al.* Discovery of novel inhibitors of human galactokinase by virtual screening. *J Comput Aided Mol Des* **33**, 405–417 (2019).

34. Seifert, M. H. J., Wolf, K. & Vitt, D. Virtual high-throughput in silico screening. *Biosilico* **1**, 143–149 (2003).

35. *Virtual Screening: An Alternative or Complement to High Throughput Screening?: Proceedings of the Workshop 'New Approaches in Drug Design and Discovery', special topic 'Virtual Screening', Schloß Rauischholzhausen, Germany, March 15–18, 1999*. (Springer Netherlands, 2002). doi:10.1007/0-306-46883-2.

36. Compound Management. *National Center for Advancing Translational Sciences* https://ncats.nih.gov/preclinical/core/compound (2015).

37. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204-212 (2015).

38. h11g11-bg - H11G11-BG protein - uncultured bacterium - h11g11-bg gene & protein. https://www.uniprot.org/uniprot/D5GU71.

39. Molecular Operating Environment (MOE) | MOEsaic | PSILO. https://www.chemcomp.com/Products.htm.

40. BLAST: Basic Local Alignment Search Tool. https://blast.ncbi.nlm.nih.gov/Blast.cgi.
41. RCSB PDB: Homepage. https://www.rcsb.org/.
42. RCSB PDB - 6MVH: Crystal structure of FMN-binding beta-glucuronidase from Roseburia hominis. https://www.rcsb.org/structure/6MVH.
43. Xiang, Z. Advances in Homology Protein Structure Modeling. *Curr Protein Pept Sci* **7**, 217–227 (2006).
44. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules | Journal of the American Chemical Society. https://pubs.acs.org/doi/abs/10.1021/ja00124a002.
45. Lee, J., Wu, S. & Zhang, Y. Ab Initio Protein Structure Prediction. in *From Protein Structure to Function with Bioinformatics* (ed. Rigden, D. J.) 3–25 (Springer Netherlands, 2009). doi:10.1007/978-1-4020-9058-5_1.
46. Peng, J. & Xu, J. A multiple-template approach to protein threading. *Proteins: Structure, Function, and Bioinformatics* **79**, 1930–1939 (2011).
47. Haddad, Y., Adam, V. & Heger, Z. Ten quick tips for homology modeling of high-resolution protein 3D structures. *PLOS Computational Biology* **16**, e1007449 (2020).
48. I-TASSER server for protein structure and function prediction. https://zhanglab.ccmb.med.umich.edu/I-TASSER/.
49. RCSB PDB - 6MVG: Crystal structure of FMN-binding beta-glucuronidase from Ruminococcus gnavus. https://www.rcsb.org/structure/6MVG.
50. Yang, J. & Zhang, Y. Protein Structure and Function Prediction Using I-TASSER. *Curr Protoc Bioinformatics* **52**, 5.8.1-5.815 (2015).

51. RCSB PDB - 3K4D: Crystal structure of E. coli beta-glucuronidase with the glucaro-d-lactam inhibitor bound. https://www.rcsb.org/structure/3K4D.
52. RCSB PDB - 5Z1A: The crystal structure of Bacteroides fragilis beta-glucuronidase in complex with uronic isofagomine. https://www.rcsb.org/structure/5Z1A.
53. RCSB PDB - 5CZK: Structure of E. coli beta-glucuronidase bound with a novel, potent inhibitor 1-((6,8-dimethyl-2-oxo-1,2-dihydroquinolin-3-yl)methyl)-1-(2-hydroxyethyl)-3-(4-hydroxyphenyl)thiourea. https://www.rcsb.org/structure/5czk.
54. UCSF Chimera Home Page. https://www.cgl.ucsf.edu/chimera/.
55. OEDocking Software | Molecular Docking Tools | Fred Docking. https://www.eyesopen.com/oedocking.
56. NIH HPC Systems. https://hpc.nih.gov/systems/.
57. Ahmad, S., Hughes, M. A., Yeh, L.-A. & Scott, J. E. Potential repurposing of known drugs as potent bacterial β-glucuronidase inhibitors. *J Biomol Screen* **17**, 957–965 (2012).
58. Blázquez, M. Quantitative GUS Activity Assay in Intact Plant Tissue. *CSH Protoc* **2007**, pdb.prot4688 (2007).
59. 4-<WBR>Methylumbelliferyl-β-<SC>D</SC>-glucuronide hydrate M9130. *881005-91-0* https://www.sigmaaldrich.com/catalog/product/sigma/m9130.
60. Perkin Elmer. *ViewLux™ Picture This*.
61. Zhang, J.H., Chung, T.D. & Oldenburg. K.R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* **4**, 67–73 (1999).
62. Hughes, J. P., Rees, S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).

63. Prism Updates - GraphPad. https://www.graphpad.com/updates/.

64. Li, J. *et al.* RUC-4: A Novel αIIbβ3 Antagonist for Pre-hospital Therapy of Myocardial Infarction. *Arterioscler Thromb Vasc Biol* **34**, 2321–2329 (2014).

65. Vootukuri, S. *et al.* Preclinical studies of RUC-4, a novel platelet αIIbβ3 antagonist, in non-human primates and with human platelets. *J Clin Transl Sci* **3**, 65–74 (2019).

66. CeleCor Therapeutics Announces Positive First-in-Human Data of RUC-4, a Novel Subcutaneous Platelet GPIIb/IIIa Inhibitor, Presented at TCT 2019 | Business Wire. https://www.businesswire.com/news/home/20190925005021/en/CeleCor-Therapeutics-Announces-Positive-First-in-Human-Data-RUC-4.

67. A Randomized Phase 1 Dose-Escalation Study of Subcutaneously(SC) Administered RUC-4 - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT03844191.

68. A Phase 2 Open Label Study to Assess the PK/PD Properties of RUC-4 in Patients With a ST-elevation Myocardial Infarction - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT04284995.

69. COVID-19–associated coagulopathy. https://www.the-hospitalist.org/hospitalist/article/221875/coronavirus-updates/covid-19-associated-coagulopathy.

70. Becker, R. C. COVID-19 update: Covid-19-associated coagulopathy. *J Thromb Thrombolysis* 1–14 (2020) doi:10.1007/s11239-020-02134-3.

71. Connors, J. M. & Levy, J. H. COVID-19 and its implications for thrombosis and anticoagulation. *Blood* **135**, 2033–2040 (2020).

72. Assay Development & Screening. *National Center for Advancing Translational Sciences* https://ncats.nih.gov/preclinical/drugdev/assay (2015).

73. Dahlin, J. L. & Walters, M. A. The essential roles of chemistry in high-throughput screening triage. *Future Medicinal Chemistry* **6**, 1265–1290 (2014).

74. Taylor, M. R. *et al.* Vancomycin relieves mycophenolate mofetil–induced gastrointestinal toxicity by eliminating gut bacterial β-glucuronidase activity. *Science Advances* **5**, eaax2358 (2019).

75. Saitta, K. S. *et al.* Bacterial β-glucuronidase inhibition protects mice against enteropathy induced by indomethacin, ketoprofen or diclofenac: mode of action and pharmacokinetics. *Xenobiotica* **44**, 28–35 (2014).

76. Bhatt, A. P. *et al.* Targeted inhibition of gut bacterial β-glucuronidase activity enhances anticancer drug efficacy. *PNAS* **117**, 7374–7381 (2020).

77. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians* **70**, 7–30 (2020).

APPENDIX A

SUPPLEMENTAL SCREENING DATA AND STRUCTURE FILES

To promote the reproducibility of the work we present in this thesis, I have created an online repository with our primary sequence of H11G11-BG, a Ramachandran plot of my MOE model, my template search and threading results from I-TASSER, and my MOE and I-TASSER homology models. This database also includes the top 500 vHTS results from our *holo* GUS screen and relevant screening data for all 69 bioactive compounds we identified from our biochemical validation assay.

At the request of Symberix, Inc., the industrial co-sponsor of our research, we are unable to release our data repository publicly at this time. Requests for the data elements above may be emailed to the author at anup.p.challa@vanderbilt.edu and will be reviewed by Symberix, Inc. for feasibility of fulfillment.

Upon public release of our online repository, I encourage readers interested in the above information to visit this resource. We also plan to deposit our homology models in a community-wide protein structure databank and hope our readers find these data helpful in replicating the reseach in this thesis.