

University of Memphis

University of Memphis Digital Commons

Electronic Theses and Dissertations

2020

Computational Prediction and Experimental Validation of ADMET Properties for Potential Therapeutics

Keri Danielle Hannie

Follow this and additional works at: <https://digitalcommons.memphis.edu/etd>

Recommended Citation

Hannie, Keri Danielle, "Computational Prediction and Experimental Validation of ADMET Properties for Potential Therapeutics" (2020). *Electronic Theses and Dissertations*. 2570.
<https://digitalcommons.memphis.edu/etd/2570>

This Dissertation is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khhgerty@memphis.edu.

COMPUTATIONAL PREDICTION AND EXPERIMENTAL VALIDATION OF
ADMET PROPERTIES FOR POTENTIAL THERAPEUTICS

by

Keri Danielle Hannie

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Chemistry

The University of Memphis

May 2020

Dedication

This is dedicated to my mother, Carla Hannie, who unfortunately passed away my last year in graduate school. She always encouraged me to follow my dreams and I know she is looking down at this accomplishment with pride. Without her constant support, I would not have gone to my school of choice and gotten to where I am today.

To my nephew, Jacen Steber, let this be a lesson that whatever you wish for may come true with a little luck and a lot of determination. Perseverance is key.

Acknowledgements

Portions of the work presented in Chapter 3 were made possible by funding from the UM CAS Research Instrumentation Initiative, which allowed the Department of Chemistry to purchase the FlexStation 3 and Tycho NT.6 instruments. I want to thank the Chemical Computing Group for providing educational license access to the software package Molecular Operating Environment, which was used for the development of the QSAR models mentioned herein. I would like to thank the High-Performance Computing at the University of Memphis for giving me access to perform the conformational searches described herein. I would also like to thank Dr. Truc Chi Pham for sharing her vast knowledge and expertise of mass spectrometry with me throughout my studies, and Dr. Kristie Ruddick for giving me advice and support through many of the tasks I had to perform here. I would like to thank the Department of Chemistry and all of the friends I made throughout my time here.

I would also like to thank my committee members: *Dr. Daniel Baker, Dr. Abby Parrill, Dr. Judith Cole, Dr. Paul Simone Jr., and Dr. Charles Garner* for all their feedback and guidance through this process. I especially want to thank my committee chair, *Dr. Baker*, for allowing me to join his research group and guiding me toward the light at the end of the tunnel. Here's to hoping it's not a train after all, but the light of fulfillment of this degree. I would also like to thank my honorary secondary advisor, *Dr. Parrill*, for your constant advice and help throughout this whole process.

I would like to thank my father, *Rodney Hannie*, for always believing in me and encouraging me to follow my dreams. You always knew when I needed to let off some steam from my frustrations when my research was not cooperating, and knowingly got me angry at you

so I could do so. Thank you. To my sister, *Jordan Steber*, thank you for always being a shoulder to lean on when times got rough and I wanted to quit.

Preface

This dissertation provides an account of efforts to develop computational tools and corresponding experimental methods to inform and prioritize ligand discovery efforts for future researchers within this lab group and beyond. These tools focus on properties of common orally bioavailable drugs, namely absorption and distribution, which comprise a portion of the critical ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties.

Once ingested, oral drugs are first absorbed across the intestinal barrier before they are transported to the liver, the site of potential phase I-III metabolism, and then are distributed throughout the body in the circulation, where most drugs bind to serum proteins. Absorption is typically measured and reported as human intestinal absorption (HIA), but these measurements can be problematic for many reasons, so most researchers use experiments in model cell lines to predict HIA. One of the more common model cell lines is Caco-2, a colonic cell line that forms monolayers where permeability across this monolayer can be measured as a prediction of this absorption. One of the most abundant blood serum proteins that drugs bind to for distribution is human serum albumin (HSA), however, a cheaper alternative that scientists typically use in experiments is bovine serum albumin (BSA), which is used in this work.

In chapter 1, we provide an overview of the drug development process, the importance of ADMET properties in drug/ligand discovery efforts, as well as examples of projects our extended research group is focused on for ligand discovery efforts. Chapter 2 discusses the process used to develop computational models to predict HIA, Caco-2 permeability, and protein binding of small molecule drugs, as well as analysis of compounds predicted incorrectly in these models. Chapter 3 discusses method development efforts for experimental validation of computational predictions to determine protein binding of multiple small molecule drugs.

Chapter 4 summarizes results discussed in chapters 2 and 3, as well as highlighting potential future directions of this project and proposing initial targets to be used in the models and experiments outlined therein. The appendix at the end of this work includes supplementary information excluded from chapter 2 for brevity.

Abstract

The drug development process in the United States is an expensive and lengthy process, usually taking a decade or more to gain approval for a drug candidate. The majority of proposed, early stage therapeutics fail, even though the typical process narrows from hundreds or thousands of small molecules down to one late stage candidate. One reason for failure is due to the drug's poor or unexpected absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties. Researchers attempt to predict ADMET properties as a way to help prioritize compounds for lead development to minimize expense and time. It was the overall goal of this project to further the prediction of two ADMET properties (absorption and distribution) through the development and application of quantitative structure-activity (QSAR) relationship computational models predicting human intestinal absorption (HIA), Caco-2 permeability (*in vivo* & *in vitro* measurements of absorption), and protein binding (measurement of distribution). These combined models would then be paired with additional experimental methods to help prioritize compounds for future ligand discovery efforts in our lab group and for our collaborators.

Five computational QSAR models for each of these three properties were created using different molecular descriptor types and solvation models in an effort to examine which approach resulted in optimal performance. The model development process and validation stages of these QSAR models is outlined herein, along with analysis and discussion of commonly mispredicted compounds. Performance was similar across all models (independent of the molecular descriptor used and the solvation models applied). Future efforts at model development will depend on the size of the dataset to be analyzed. If the dataset is small, the i3D-Born solvation models will be used because these models better represent physiological conditions and

performed slightly better than the other models. However, if the dataset is large, the 2D descriptor models will be used as these models do not require that a time and resource-intensive conformational search be performed and because it performed nearly as well as the i3D-Born solvation models. There were no common structural features consistently found associated with mispredicted structures. As such we are unable, at this time to pinpoint classes of compounds to avoid in future efforts

The experimental methods outlined in this work focused on developing methods to determine protein binding, specifically determining a fast, inexpensive workflow to classify the difference between high and low protein binding small molecules. Two techniques were used to determine protein binding of small molecules to bovine serum albumin (BSA): fluorescence polarization (FP) competition, and Nano Differential Scanning Fluorimetry (NanoDSF). FP assays quantifies the change in polarization of a target fluorophore between its protein bound and free states, an equilibrium that can be impacted by the presence of small molecule competitors. This method can be performed in a quantitative manner, but it also requires more time and more expensive and specialized instrumentation. In contrast, NanoDSF determines the melting temperature of BSA in the presence (higher) or in the absence (lower) small molecules by determining the intrinsic fluorescence of tryptophan and tyrosine residues while applying a temperature gradient. This method is qualitative, at least in our approach, but is very fast and requires much less expensive instrumentation. In our hands both techniques were successful in distinguishing differences between small molecules exhibiting low and high BSA binding.

In summary, this project was successful in that we 1) developed computational tools capable of correctly predicting ADMET properties including HIA, Caco-2 permeability, and protein binding and 2) developed experimental workflows to quantitatively and qualitatively

separate small molecules into low and high affinity BSA binders. With these *in silico* models and *in vitro* methods established, future research in our group and with our collaborators can make use of these tools to help prioritize compounds in ligand/ inhibitor discovery efforts.

Table of Contents

| Chapter | Page |
|--|------|
| List of Tables | xiv |
| List of Figures | xvi |
| List of Abbreviations and Acronyms | xx |
| 1. Current Challenges in the Drug Development Process | 1 |
| Introduction | 1 |
| Attrition Rates of Potential Therapeutics | 3 |
| Importance of ADMET Properties in Drug Action | 4 |
| Absorption | 5 |
| Distribution | 6 |
| Metabolism | 6 |
| Excretion | 7 |
| Toxicity | 8 |
| Ligand/Inhibitor Discovery Efforts | 9 |
| G Protein-Coupled Receptors | 9 |
| Autotaxin Inhibitors and Lysophosphatidic Acid Ligands | 10 |
| Sphingosine Kinase Inhibitors and Sphingosine-1-Phosphate Ligands | 11 |
| Organization of this Dissertation | 12 |
| 2. Computational Prediction of ADMET Properties: Absorption and Distribution | 14 |
| Introduction | 14 |
| ADMET Property: Absorption | 15 |
| ADMET Property: Distribution | 16 |
| Quantitative Structure-Activity Relationships | 16 |
| Methods | 19 |
| Database Compilation | 19 |

| | |
|--|----|
| Activity Categorization | 19 |
| Training and Test Set Selection | 20 |
| HIA and Caco-2 Permeability Training and Test Sets | 21 |
| Percent Protein Binding Training and Test Sets | 30 |
| Chemical Diversity Measures | 43 |
| QSAR Model Development, Internal Validation, and External Validation | 44 |
| Examination of Commonly Mispredicted Compounds | 47 |
| Results and Discussion | 48 |
| Dataset Diversity | 48 |
| Comparison of 5 Descriptor Model Types | 49 |
| Commonly Mispredicted Compounds | 57 |
| Conclusions | 60 |
| 3. Method Development of Protein Binding Assays | 62 |
| Introduction | 62 |
| Fluorescence Polarization | 63 |
| Nano Differential Scanning Fluorimetry | 66 |
| Methods | 68 |
| Fluorescence Polarization | 68 |
| Fluorescein Fluorescence Polarization Experiments | 69 |
| Absorbance and Emission Measurements of Fluorescein | 69 |
| Fluorescence Polarization Measurements of Fluorescein | 71 |
| Fluorescein Binding to Varying Concentrations of BSA | 73 |
| Competition Assays of BSA Binding (Fluorescein and Drugs) | 73 |
| BODIPY-LPA Fluorescence Polarization Experiments | 76 |
| Absorbance and Emission Measurements of BODIPY-LPA | 76 |
| Fluorescence Polarization Measurements of BODIPY-LPA | 78 |
| BODIPY-LPA Binding to Varying Concentrations of BSA | 79 |
| Competition Assays of BSA Binding (BODIPY-LPA and Drugs) | 80 |
| Nano Differential Scanning Fluorimetry | 81 |

| | |
|---|-----|
| Control Assays | 82 |
| Binding Assays | 82 |
| Results and Discussion | 83 |
| Fluorescence Polarization | 83 |
| Fluorescein Experiments | 83 |
| Absorbance and Emission Spectra for Fluorescein | 83 |
| Fluorescence Polarization of Fluorescein | 84 |
| Fluorescein Binding to Varying Concentrations of BSA | 85 |
| Competition Binding Assays of BSA, Fluorescein, and Drugs | 86 |
| BODIPY-LPA Experiments | 91 |
| Absorbance and Emission Spectra for BODIPY-LPA | 91 |
| Fluorescence Polarization of BODIPY-LPA | 92 |
| BODIPY-LPA Binding to Varying Concentrations of BSA | 93 |
| Competition Binding Assays of BSA, BODIPY-LPA, & Drugs | 95 |
| Nano Differential Scanning Fluorimetry | 100 |
| Control Assays | 100 |
| Binding Assays | 101 |
| Conclusions | 104 |
| 4. Conclusions and Future Approaches | 106 |
| Summary of Major Points from Computational Approaches | 106 |
| Model Selection and Future Application | 106 |
| Examination of Mispredicted Compounds | 107 |
| Summary of Major Points from Experimental Approaches | 108 |
| Fluorescence Polarization Experiments | 108 |
| NanoDSF Experiments | 109 |
| Recommendations for Future Experiments | 109 |
| Future Approaches | 110 |
| Additional Computational Work to Expand Established Methods | 110 |
| Additional Computational Methods to be Applied | 112 |

| | |
|--|-----|
| Predicting Cytochrome P450 Interactions to Determine Potential Metabolism Reactions | 112 |
| Experimental Work Utilizing Established Methods | 113 |
| Experimental Determination of Protein Binding | 113 |
| Additional Experimental Work | 113 |
| Method Development and Experimental Determination of Caco-2 Permeability | 113 |
| Method Development and Experimental Determination of Metabolism | 114 |
| References | 116 |
| Appendix | 130 |

List of Tables

| Table | Page |
|--|-------------|
| 1. Training set 1 compounds and corresponding HIA, Caco-2 permeability, and protein binding data..... | 23 |
| 2. Test set 1 compounds and the corresponding HIA, Caco-2 permeability, and percent protein binding data..... | 26 |
| 3. Training set 2 and the corresponding HIA and Caco-2 permeability data..... | 28 |
| 4. Test set 2 compounds and the corresponding HIA and Caco-2 permeability data..... | 30 |
| 5. Training set 3 compounds selected for % protein binding QSAR models..... | 31 |
| 6. Test set 3 compounds selected for % protein binding QSAR models..... | 32 |
| 7. Structures of all compounds selected to build QSAR models..... | 33 |
| 8. Comparison of all five models made with different descriptor types and solvation methods to predict HIA..... | 55 |
| 9. Comparison of all five models made with different descriptor types and solvation methods to predict Caco-2 permeability..... | 55 |
| 10. Comparison of all five models made with different descriptor types and solvation methods to predict protein binding..... | 56 |
| 11. Drugs selected to be examined using the FlexStation 3 and Tycho NT.6. Protein binding, HIA, Caco-2 permeability, & structures shown..... | 68 |

| | |
|--|----|
| 12. FlexStation 3 protocol created in the SoftMax Pro software used for absorbance measurements of fluorescein..... | 70 |
| 13. FlexStation 3 protocol created in the SoftMax Pro software used for emission measurements of fluorescein..... | 71 |
| 14. FlexStation 3 protocol created in the SoftMax Pro software used for fluorescence polarization measurements of fluorescein..... | 72 |
| 15. FlexStation 3 protocol created in the SoftMax Pro software used for absorbance measurements of BODIPY-LPA..... | 76 |
| 16. FlexStation 3 protocol created in the SoftMax Pro software used for emission measurements of BODIPY-LPA..... | 77 |
| 17. FlexStation 3 protocol created in the SoftMax Pro software used for fluorescence polarization measurements of BODIPY-LPA..... | 79 |

List of Figures

| Figure | Page |
|---|------|
| 1. The drug development process used by the U.S. FDA..... | 2 |
| 2. General structure of a G Protein-Coupled Receptor (GPCR) embedded in a lipid bilayer..... | 10 |
| 3. Autotaxin (ATX) catalyzed hydrolysis of lysophosphatidylcholine (LPC) into lysophosphatidic acid (LPA) and choline..... | 11 |
| 4. Sphingosine kinase (SK) catalyzed transfer of the γ -phosphate from ATP to sphingosine (SPH) to form sphingosine-1-phosphate and ADP..... | 11 |
| 5. Workflow of the development, validation, and application of ADMET QSAR models..... | 45 |
| 6. Representative histogram of all compound chemical diversity calculated from MACCS keys using Tanimoto coefficients..... | 48 |
| 7. Model comparisons at the model development stage and internal validation stage for human intestinal absorption data..... | 49 |
| 8. Model comparison at the external validation stage for human intestinal absorption data..... | 50 |
| 9. Model comparisons at the model development stage and internal validation stage for Caco-2 permeability data..... | 51 |
| 10. Model comparison at the external validation stage for Caco-2 permeability data..... | 52 |
| 11. Model comparisons at the model development stage and internal validation stage for protein binding data..... | 53 |

| | |
|---|----|
| 12. Model comparison at the external validation stage for protein binding data..... | 54 |
| 13. Structure of hydralazine..... | 58 |
| 14. Histogram of the commonly mispredicted compounds within the HIA QSAR models..... | 58 |
| 15. Histogram of the commonly mispredicted compounds within the Caco-2 permeability QSAR models..... | 59 |
| 16. Structure of tranexamic acid. The only compound commonly mispredicted in all the HIA and Caco-2 permeability QSAR models..... | 60 |
| 17. Schematic representation of fluorescence polarization..... | 64 |
| 18. Structure of fluorescein sodium salt used in these experiments..... | 66 |
| 19. Structure of TopFluor LPA (BODIPY-LPA) used in these experiments..... | 66 |
| 20. Schematic representation of Nano Differential Scanning Fluorimetry to determine protein binding..... | 67 |
| 21. Absorbance and Emission spectra of fluorescein sodium salt and buffer (1X PBS, pH 7.4).. | 83 |
| 22. Perpendicular and parallel emission spectrum of dose response of fluorescein in buffer (1X PBS, pH 7.4)..... | 85 |
| 23. Fluorescence polarization (milliPolarization) of varying BSA concentrations (0, 1.0, 10.0, & 100.0 μ M) in the presence of 2.5 μ M fluorescein..... | 86 |
| 24. Average milliPolarization of atenolol dose response to 10.0 μ M BSA and 2.5 μ M fluorescein..... | 87 |

| | |
|---|-----|
| 25. Average milliPolarization of hydrochlorothiazide dose response to 10.0 μM BSA and 2.5 μM fluorescein..... | 88 |
| 26. Average milliPolarization of naproxen dose response to 10.0 μM BSA and 2.5 μM fluorescein..... | 89 |
| 27. Average milliPolarization of ibuprofen dose response to 10.0 μM BSA and 2.5 μM fluorescein..... | 90 |
| 28. Absorbance and Emission spectra of BODIPY-LPA and buffer (1X PBS, pH 7.4)..... | 91 |
| 29. Perpendicular and parallel emission spectrum of dose response of BODIPY-LPA in buffer (1X PBS)..... | 93 |
| 30. Fluorescence polarization of varying BSA concentrations (0, 1.0, 10.0, & 100.0 μM) in the presence of 1.0 μM BODIPY-LPA..... | 94 |
| 31. Average milliPolarization of atenolol dose response to 10.0 μM BSA and 1.0 μM BODIPY-LPA..... | 95 |
| 32. Average milliPolarization of hydrochlorothiazide dose response to 10.0 μM BSA and 1.0 μM BODIPY-LPA..... | 96 |
| 33. Average milliPolarization of naproxen dose response to 10.0 μM BSA and 1.0 μM BODIPY-LPA..... | 98 |
| 34. Average milliPolarization of ibuprofen dose response to 10.0 μM BSA and 1.0 μM BODIPY-LPA..... | 99 |
| 35. Average change in ratio (350nm/330nm) as a function of temperature for each compound... | 101 |
| 36. Average T_m of 10.0 μM BSA in the presence of varying concentrations of the drug atenolol..... | 102 |

| | |
|--|-----|
| 37. Average T_m of 10.0 μM BSA in the presence of varying concentrations of the drug hydrochlorothiazide..... | 103 |
| 38. Average T_m of 10.0 μM BSA in the presence of varying concentrations of the drug ibuprofen..... | 104 |
| 39. Structures of seven initial targets..... | 111 |

List of Abbreviations and Acronyms

2D: Two-Dimensional

3D: Three-Dimensional

ADMET: Absorption, Distribution, Metabolism, Excretion, & Toxicity

ATCC: American Type Culture Collection

ATP: Adenosine Triphosphate

ATX: Autotaxin

BBB: Blood-Brain Barrier

BLA: Biologics License Application

BSA: Bovine Serum Albumin

BODIPY-LPA: BODIPY-Labeled Lysophosphatidic Acid

BNN: Bayesian Neural Network

CART: Classification and Regression Trees

CYP: Cytochrome

C Field: Contingency Coefficient

°C: Degrees Celsius

DMSO: Dimethyl Sulfoxide

DPBS: Gibco's Dulbeccos's 1X Phosphate Buffered Saline

EL: Extracellular Loop

FDA: Food and Drug Administration

FP: Fluorescence Polarization

F_{||}: Parallel Fluorescence

F_⊥: Perpendicular Fluorescence

GPCR: G Protein-Coupled Receptor

GPR52: G Protein-Coupled Receptor 52

GPR88: G Protein-Coupled Receptor 88

HIA: Human Intestinal Absorption

HPC: High-Performance Computer

HSA: Human Serum Albumin

HTS: High-Throughput Screening

IADR: Idiosyncratic Adverse Drug Reaction

IAM: Immobilized Artificial Membrane

IL: Intracellular Loop

LOO: Leave One Out

logP: Octanol/Water Partition Coefficient

LPA: Lysophosphatidic Acid

LPA₁₋₆: Lysophosphatidic Acid Receptors 1-6

LPC: Lysophosphatidylcholine

MACCS: Molecular ACCess System

MDCK: Madin-Darby Canine Kidney

mP: milliPolarization

mM: millimolar Concentration

μM: micromolar Concentration

MOE: Molecular Operating Environment

MgCl₂: Magnesium Chloride

MMFF94x: Merck Molecular ForceField 94x

NADP⁻: reduced Nicotinamide Adenine Dinucleotide Phosphate

NADP⁺: Nicotinamide Adenine Dinucleotide Phosphate

NanoDSF: Nano Differential Scanning Fluorimetry

NDA: New Drug Application

P: Polarization

PAB: Permeability Assay Buffer

PAMPA: Parallel Artificial Membrane Permeability Assay

P_{app}: Apparent Permeability

PB: Protein Binding

PBS: Phosphate Buffered Saline

PD: Pharmacodynamics

PDB: Protein Data Bank

PK: Pharmacokinetics

PLS: Partial Least Squares

PMT: Photomultiplier Tube

QSAR: Quantitative Structure-Activity Relationship

QSPR: Quantitative Structure-Property Relationship

RFU: Relative Fluorescence Units

R Field: Correlation Coefficient

RMs: Reactive Metabolites

SAR: Structure-Activity Relationship

SK: Sphingosine Kinase

SK1: Sphingosine Kinase 1

SPH: Sphingosine

S1P: Sphingosine-1-Phosphate

S1P₁₋₅: Sphingosine-1-Phosphate Receptors 1-5

SVM: Support Vector Machine

T_i : Inflection Temperature

T_m : Melting Temperature

TM: Transmembrane

UGT: UDP-Glucuronosyltransferase

U Field: Entropic Uncertainty

UTHSC: University of Tennessee Health Science Center

V Field: Cramer's V

Chapter 1

Current Challenges in the Drug Development Process

Introduction

The process of developing a Food and Drug Administration (FDA) approved drug is long and extremely expensive. This process is composed of five main steps and can take more than a decade in most cases.¹ Figure 1 shows the general steps required for a typical drug development process. Recently, it was found that it costs, on average, 2.6 billion US dollars to develop a new therapeutic.^{2,3} The attrition rate for drug candidates is estimated to be up to 96%. As such only 4% of initially identified drug candidates will ultimately be approved as therapeutics.¹ This significantly over estimates the attrition associated with very early small molecules identified as receptor ligands or enzyme inhibitors. Thus, strategies need to be implemented as early as possible in workflows to eliminate small molecules likely to fail in subsequent drug development efforts. One such approach is to predict, measure, and understand the absorption, distribution, metabolism, excretion, and toxicity (ADMET)⁴ properties of potential drugs.

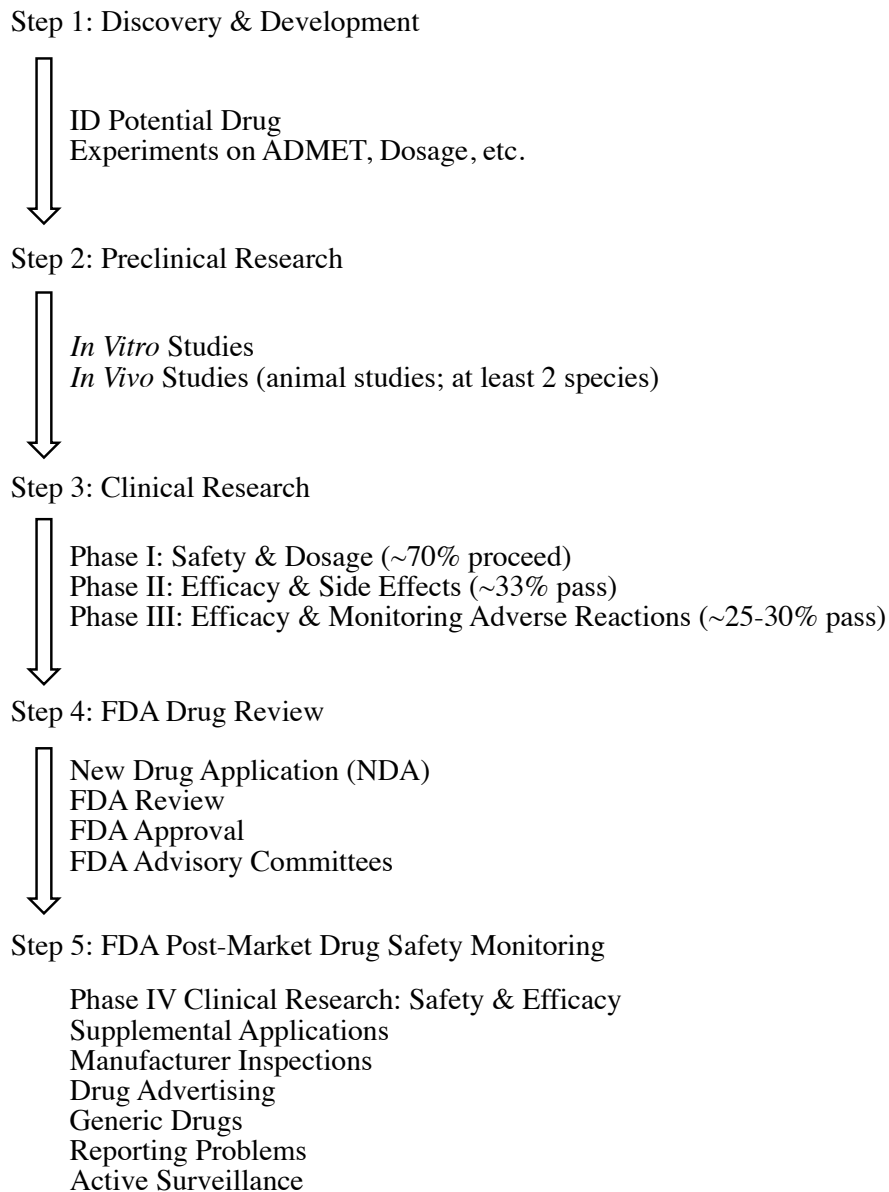


Figure 1. The drug development process used by the U.S. FDA.¹ In step 1, a potential drug is identified, and experiments are performed by the developer to potentially determine ADMET properties and dosages. Step 2 includes *in vitro* and *in vivo* studies on the drug before testing occurs in humans. If the drug passes these steps, the clinical trials of step 3 occur. The clinical trials show increasing attrition in each phase, with ~70%, 33%, and 25-30% proceeding from phase I, II, and III, respectively. Once phase III has been successfully passed, a new drug application can occur, after which the drug is reviewed by the FDA for approval. After drug approval occurs, monitoring by the FDA continues in step 5.

Attrition Rates of Potential Therapeutics

The high attrition rate of potential therapeutics (96%) has led several researchers to investigate the reason(s) for failure. In 2015, Waring *et al.*, examined the attrition rate associated with the efforts of four major pharmaceutical companies (AstraZeneca, Eli Lilly and Company, GlaxoSmithKline, and Pfizer) from 2000-2010.⁵ Of the 808 compounds listed from all four companies with the highest phase of development provided, a breakdown of progression from nomination to each phase was then examined.⁵ This summary consisted of 422 total compounds in the nomination stage (356 failed at this stage, 60 moved into phase I, & 6 were not specified), phase I included 231 compounds (157 failed, 71 moved into phase II, & 3 were not specified), phase II was comprised of 145 compounds (89 failed, 54 moved into phase III, & 2 were not specified), phase III included 8 compounds (2 failed, & 6 moved into phase IV), and finally phase IV consisted of 2 compounds, both of which passed.⁵ This summary, across all stages has an overall failure rate of 76%, with 615 overall of the 808 compounds not progressing if you assume that all unspecified compounds failed.

In 2010, DiMasi *et al.* examined the success rate of investigational drugs from the 50 largest pharmaceutical firms during the period 1993-2004.⁶ This study examined a total of 1,738 compounds and calculated the success rate to be 7.9% (based on approved compounds (138) at the time of publication). In addition, this work calculated a maximum possible success rate of 32.2% (assuming all proposed compounds (422) did not fail).⁶

In 2018, Takebe *et al.*, investigated the success rate of 798 academic projects in the drug development industry from 1991-2010.⁷ In this work 31.8% of candidates passed from pre-clinical-phase I, 75.1% passed from phase I-phase II, 50% passed from phase II-phase III, 58.6%

passed from phase III-BLA/NDA review, and 87.5% passed from BLA/NDA review-approval; with an overall likelihood of approval rate of 19.3%.⁷

These high attrition rates give medicinal chemists reason to try to predict ADMET properties and to identify additional reasons to remove small molecules from drug discovery pipelines before investing significant time and resources. Ways to correctly predict undesirable properties or uncover potential side effects would have the potential to save time and resources in the development of novel small molecules as drug candidates, which is the overall goal of the work described herein.

Importance of ADMET Properties in Drug Actions

Unfavorable ADMET properties are widely recognized as reasons that many potential candidates fail in the early stages of the drug development process.^{8,9} Understanding and accurately predicting ADMET properties is an important goal, however, it is often difficult to evaluate compounds identified in high throughput screening (HTS) due to a scarcity of measured ADMET data.^{4,10} Medicinal chemists turn to predictions of ADMET properties to help estimate how a candidate will be absorbed across cell membranes, how it will be distributed in the bloodstream, what the effects of Phase I-III metabolism will be, how it will be excreted from the body, and whether it will be toxic to a patient. These ADMET properties all depend on complex processes, which often are interconnected. Typically, medicinal chemists predict these properties using *in silico* methods, usually through linear regression models, neural networks, regression trees, or support vector machines (SVM).¹¹

Absorption

Absorption is typically discussed in terms of orally bioavailable drugs, however, it can also mean other administration routes including nasal, intravenous, intramuscular, oral inhalation, rectal, vaginal, or subcutaneous.¹² Oral drugs must be absorbed across the stomach or the intestinal barrier, and this absorption can be affected by conditions at the site of absorption, drug concentrations, the pH of the intestinal contents, the circulation at the site of absorption, and the area of the surface the drug absorbs through.¹³ Small lipophilic compounds absorb more easily than large compounds, and unionized compounds absorb into the body more easily than ionized compounds.¹⁴

Measurements to determine physicochemical properties that contribute to absorption, like solubility and passive intestinal absorption, have been performed previously using solubility assays, immobilized artificial membranes (IAMs), and parallel artificial membrane permeability assays (PAMPA).¹⁵⁻¹⁷ Unidirectional cell permeability assays have used several cell lines, such as Caco-2 (human colorectal adenocarcinoma cells), Madin-Darby canine kidney (MDCK), and HT29 (human colorectal adenocarcinoma cells), to determine a model system that reflects human intestinal absorption (HIA).¹⁵⁻¹⁷ Models predicting absorption typically predict Caco-2 permeability and HIA based on regression techniques, and are often called quantitative structure-activity relationship (QSAR) or quantitative structure-property (QSPR) models.¹⁸⁻²⁰ The introduction to chapter 2 explains why these already developed models were not used in this project to predict absorption.

Distribution

Distribution of drugs occurs through the bloodstream, which is impacted by the amount of blood flow in certain areas (well supplied areas receive the drug more quickly), as well as how much of the drug is bound to plasma proteins (acidic drugs bind to albumin, basic drugs bind to α_1 -acid glycoprotein).¹³ Physical properties of the drug, such as aqueous solubility and partition coefficients, impact how they are distributed throughout the body.¹² The amount of drug bound to serum proteins is determined by drug concentration, the number of binding sites on the protein, and the drug's affinity for the binding site.¹³ Drugs bound to serum proteins typically do not reach their receptors in a timely manner and are usually retained in the body longer due to urine being protein-free (decreased excretion); however, most drugs are not highly plasma protein bound and reach their receptors quickly to elicit the desired interactions.¹⁴

Distribution measurements to determine blood-brain barrier (BBB) penetration can be performed using PAMPA assays similar to those determining absorption, but using BBB specific lipids, efflux assays examining calcium uptake, p-glycoprotein-ATPase assays, and protein binding assays examining human serum albumin (HSA) binding have all been performed.^{16, 17} Distribution models have been developed to predict steady-state volumes of distribution, using Bayesian neural networks (BNN), classification and regression trees (CART) and partial least squares (PLS).¹⁹ The introduction to chapter 2 explains why models previously developed were not used to predict distribution in this project.

Metabolism

Metabolism is interrelated to both excretion and toxicity; metabolism often helps the body rid itself of the drug by promoting excretion through metabolite formation, and some of the

resulting metabolites can be toxic (extremely undesirable).¹² Metabolism biotransformations usually involve phase I and phase II reactions in the liver with phase I involving oxidation, reduction, or hydrolysis, and phase II involving conjugation reactions.¹³ Metabolism creates more polar metabolites that are more readily excreted in the urine and bile; when phase I reactions do not produce a molecule with sufficient polarity, phase II reactions such as glucuronidation or sulfation can occur.¹⁴

Metabolism measurements characteristically examine metabolic stability using human or rat liver microsomes or hepatocytes.^{17, 21} Drug-drug interactions can be examined through cytochrome (CYP) P450 inhibition screens (as certain drugs are known to impact the metabolic stability of other drugs), or IC₅₀ determinations.^{15, 17, 21} QSAR models have been developed to predict CYP metabolism and UDP-Glucuronosyltransferase (UGT) metabolism.^{19, 22} Models have also been developed to predict the site of metabolism.

Excretion

Given sufficient time, the remainder of an administered drug and/ or its metabolites are excreted from the body, predominantly through the renal system (urine), but can also occur through fecal matter (bile), sweat, saliva, and tears (though these last three routes are rare).¹³ Most drugs or metabolites that are excreted via the bile enter the small intestine, where they are excreted in the feces, however, entero-hepatic cycling can occur due to bile salt recycling. This process occurs when additional intestinal metabolism results or when the drug or its metabolite(s) are reabsorbed and enter the liver once again.¹⁴ Models predicting clearance have been developed; some have predicted the log D-dependency of renal clearance, hepatic uptake,

and biliary clearance have been predicted through SAR by others, and renal clearance has also been modeled, though in this instance it was reported as a percentage of total clearance.¹⁹

Toxicity

Some parent drugs are toxic or biotransformation reactions can create toxic metabolites.¹³ Several mechanisms of toxicity can occur: on-target (mechanism-based) toxicity, hypersensitivity/immune response toxicity, off-target toxicity, bioactivation, and idiosyncratic reactions.²³ On-target, or metabolism-based, toxicity occurs when the interaction between the drug and the target produces both an efficacious response and a toxic response.²³ Hypersensitivity, or immune response toxicity, is caused when a drug or its metabolite interacts with proteins and cause antibodies to be induced.²³ Off-target toxicity occurs when the drug interacts with several targets, one with the intended response, and the other with a toxic response.²³ Bioactivation describes the generation of toxic metabolites that can form covalent adducts with macromolecules and cause a rare immune response called idiosyncratic adverse drug reactions (IADR), which only occur in 1/10³ or 1/10⁴ individuals.^{12, 23} Due to toxicity being hard to control, researchers typically try to avoid certain functional groups known to common causes such as thioureas, aromatic nitro groups, furan rings, and electrophilic functional groups when designing a new drug.¹⁴

Toxicity typically is measured using basic cytotoxicity assessments such as mitochondrial inhibition assays or genotoxicity measurements in the form of AMES assays.^{17, 21} The toxic effects of metabolites have been predicted, though this is the more difficult ADMET property to correctly anticipate.²²

Ligand/Inhibitor Discovery Efforts

Many researchers in both academic and industrial labs examine ligand- and receptor-mediated activation of cellular signaling pathways thought to be involved in various diseases and disorders as a way to develop tools for additional analysis of pathways and or diseases and in some cases to help prioritize compounds for assessment as potential therapeutics. Some of the ligands of interest to our lab and to our collaborators include autotaxin inhibitors, lysophosphatidic acid receptor ligands, sphingosine kinase inhibitors, sphingosine-1-phosphate receptor ligands, and ligands (agonists and antagonists) of G protein-coupled receptors, specifically orphan GPCR.

G Protein-Coupled Receptors

G protein-coupled receptors comprise a large family of membrane proteins involved in transmitting an extracellular signal into cells.²⁴ Of all currently available drugs, it is estimated that 30-60% act on this large family of GPCR.²⁵ GPCR involvement has been demonstrated in many biological processes including cardiovascular functions, cognitive responses, and cancer cell growth.²⁶⁻²⁸ Figure 2 shows the general structure of GPCR. Orphan GPCR are G protein-coupled receptors that have an unknown endogenous ligand. Of the ~400 non-sensory GPCR identified from human genome sequencing data,²⁹ approximately 100 are orphan GPCR and their ligands have yet to be identified (deorphaned).^{30, 31}

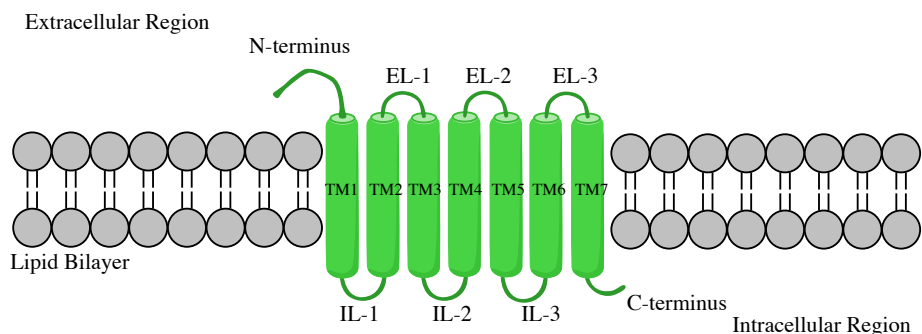


Figure 2. General structure of a G Protein-Coupled Receptor (GPCR) embedded in a lipid bilayer. Seven transmembrane spanning domains (TM1-7) make up the GPCR, with an extracellular N-terminus on TM1 and an intracellular C-terminus on TM-7. There are three extracellular loops (EL1-3) and three intracellular loops (IL1-3), with an interacting domain on IL-3 that has selectivity for a specific G protein and a segment of IL-2 that helps with efficiency of activation.

Autotaxin Inhibitors and Lysophosphatidic Acid Ligands

Autotaxin (ATX) is an ectoenzyme that catalyzes the hydrolysis of lysophosphatidylcholine (LPC) to form the bioactive lipid lysophosphatidic acid (LPA)³² as shown in Figure 3. LPA is a ligand for a series of GPCR known as LPA₁₋₆,³³ which transduce signals resulting in a variety of cellular responses including cell proliferation, migration, and prevention of apoptosis. These cellular responses have linked LPA and ATX to the pathogenesis of several human diseases including cancer, fibrotic disorders, neuropathic pain, reproductive disorders, and arthritis.³⁴ Several ATX inhibitors have been developed, including the first inhibitor in clinical trials for idiopathic pulmonary fibrosis (GLPG1690),³⁵ PAT-352,³⁶ Merck 17,³⁷ GRI 182135,^{38,39} and HA 155.⁴⁰

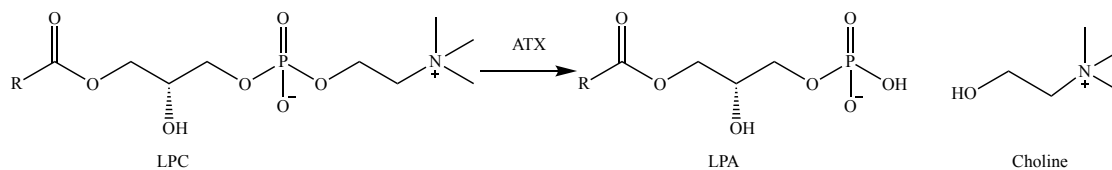


Figure 3. Autotaxin (ATX) catalyzed hydrolysis of lysophosphatidylcholine (LPC) into lysophosphatidic acid (LPA) and choline. R is a hydrocarbon chain of varying lengths.

Sphingosine Kinase Inhibitors and Sphingosine-1-Phosphate Ligands

Sphingosine kinases (SK) are enzymes that catalyze the transfer of the γ -phosphate from adenosine triphosphate (ATP) to sphingosine (SPH) to form the bioactive lipid sphingosine-1-phosphate (S1P) as shown in Figure 4. S1P is a ligand for a series of GPCR known as S1P₁₋₅ and GPR6.^{41, 42} S1P regulates several biological processes including cell motility, apoptosis, ion mobilization, cell survival, and cell proliferation.⁴²⁻⁵¹ Overexpression of the SK isoform sphingosine kinase 1 (SK1) has been linked to several cancers; in human breast cancer cells, SK1 activity regulates cancer cell proliferation and resistance to chemotherapy treatments, which has led researchers to target this signaling mechanism as a therapeutic pathway.⁵² Several SK inhibitors have been developed with this goal in mind, such as SKi,⁵³ FTY720,⁵⁴ and (*S*)-FTY720.⁵⁴

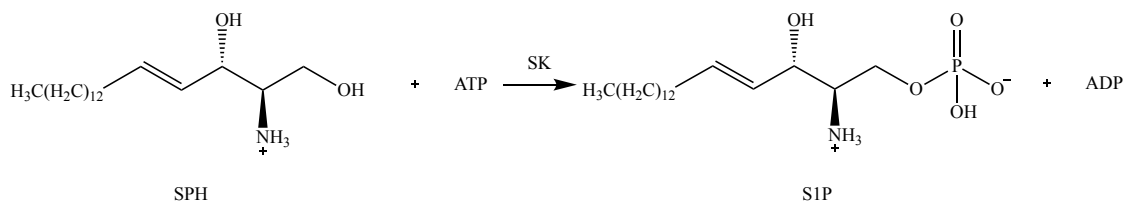


Figure 4. Sphingosine kinase (SK) catalyzed transfer of the γ -phosphate from ATP to sphingosine (SPH) to form sphingosine-1-phosphate and ADP.

Organization of this Dissertation

Chapter 2 details the development and evaluation of the performance of computational models predicting ADMET properties of drugs (absorption and distribution), specifically human intestinal absorption (HIA), Caco-2 permeability, and percent protein binding. The models outlined in chapter 2 were quantitative structure-activity relationship (QSAR) models, which have been used previously to predict ADMET properties. However, these models are typically not readily available to the public or require a significant investment to purchase them. The model development process, as well as internal and external validation procedures and results are presented in this chapter in efforts to evaluate model performance.

Chapter 3 describes the development of protein binding assays using fluorescence polarization and nano differential scanning fluorimetry to experimentally determine protein binding between bovine serum albumin (BSA) and compounds of interest. Fluorescence polarization assays have been previously used to determine protein binding; the earliest example was in 1952.⁵⁵ The fluorescence polarization methods outlined in this chapter are a semi-quantitative way to determine percent protein binding, however, these methods are costly in both time and money, as a fluorophore and specialized instrumentation are both required. Nano differential scanning fluorimetry has been used to determine protein binding using thermal shift analysis,⁵⁶ but to the best of our knowledge has not been used to determine a difference in high or low protein binding drugs, as was the case in this dissertation. In our hands, nano differential scanning fluorimetry was able to differentiate low from high protein binders in a label free, qualitative manner, using instrumentation that is significantly less expensive than that required for fluorescence polarization.

Chapter 4 provides a summary of the results presented in chapters 2 and 3 along with suggested future directions of the project. This chapter presents ways that the methods described herein can be used to help prioritize potential therapeutics.

The appendix contains supplementary information for chapter 2, including relevant tables and figures.

Chapter 2

Computational Prediction of ADMET Properties: Absorption and Distribution

Introduction

Reliable *in silico* methods that are capable of accurately predicting ADMET properties can aid in the drug development process. These predictions can be used to help eliminate compounds early in the process by identifying those targets that have undesirable ADMET properties. This ability would reduce the time and resources spent testing compounds that would likely fail at later points in the process. At the time this research project was initiated, there was a lack of availability of free or low cost ADMET prediction models. While several sources claimed free availability for downloadable software, however, I found that the websites either did not work or required computer systems that were not supported in our lab.

Examples are described herein. FAF-Drugs2, a software predicting the A, D, M, & T of ADMET, was published in 2008, and is available for download, but a Linux system is required to run it, which is a system our lab does not support, preferring either Windows or Mac software.⁵⁷ In 2012, admetSAR, which predicts all ADMET properties, was published but the website did not work when the project was initiated in 2017.⁵⁸ The developers of this software have since created a new website, but it was not available until 2019, when our models had already been developed.⁵⁹ PaDEL-DDPredictor, predicting pharmacodynamics (PD), pharmacokinetics (PK), and the T of ADMET properties, also published also in 2012, had a nonfunctional website that did not work when this project began.⁶⁰ ChemoPY, predicting structure and physicochemical properties, published in 2013, once again had a nonfunctional website at the time of this project.⁶¹

Currently, several free sources are available for use (namely the vNN webserver,⁶² and ADMETlab, that both predict all ADMET properties,⁶³ and CypReact, that predicts CYP 450 enzyme reactants⁶⁴), but they were only made available to the public after this research project began and our models were already developed. There are other software packages available, but they require licensing, like ADMET Predictor from Simulations Plus, which predicts all ADMET properties.⁶⁵

Due to the lack of availability of easily accessible tools and a desire to generate a workflow that would be more tractable to a wider group of investigators, we endeavored to generate our own models that could predict ADMET properties. The specific ADMET properties examined in this work were absorption and distribution, as they are two of the first ADMET properties that influence *in vivo* drug fate after ingestion. The *in silico* method chosen within this work is quantitative structure-activity relationship (QSAR) modeling.⁶⁶ Models used to predict these particular ADMET properties can later be used as targeted filters for known compounds of interest to our lab or to our collaborators and others, as well as filters for large compound databases to target searches toward compounds with desirable properties. These filters will help prioritize compounds of interest to be examined experimentally as potential orally bioavailable drugs.

ADMET Property: Absorption

Models for absorption prediction were constructed using training examples characterized by two types of data: human intestinal absorption and Caco-2 permeability, examples of good absorption for both HIA and Caco-2 permeability are shown in the activity categorization section of this chapter. Human intestinal absorption (HIA) is an important consideration for orally

ingested drugs due to the requirement that they be stable in the gastrointestinal tract and be able to cross the intestinal wall.⁶⁷ This absorption measurement is challenging and in many cases unethical to perform using investigational compounds, thus Caco-2 permeability was also modeled. Caco-2 is a cell line commonly used to model the intestinal barrier.⁶⁸ These cells form monolayers in culture to resemble the epithelial cells of the intestinal wall and were originally derived from human colon adenocarcinoma cells.⁶⁹ The *in vivo* data (HIA) and the *in vitro* data (Caco-2 permeability) were correlated before models were developed to ensure the suitability of the Caco-2 model system as a surrogate for predictions of HIA (details provided below).

ADMET Property: Distribution

One aspect of distribution was modeled on the basis of percent protein binding data. Protein binding measurements can help determine how much of a drug is bound to blood serum proteins and how much of the unbound drug is available to interact with target cells throughout the body, examples of protein binding values showing good distribution are shown later in the activity categorization section of this chapter.⁷⁰ Models predicting this ADMET property were developed in conjunction with the absorption models' data to narrow down compounds of interest assessment of experimental biological activity.

Quantitative Structure-Activity Relationships

The field of quantitative structure-activity relationship (QSAR) modeling was founded in 1962 by Corwin Hansch.⁷¹ This seminal paper ended a 15-year period where researchers attempted to understand structure-activity relationships (SARs) of plant growth regulators, where most researchers attempted to find a suitable Hammett relationship.⁶⁶ The Hammett relationship

is based on an equation relating reaction rates and equilibrium constants, developed by Louis P. Hammett in 1935.⁷² The Hammett equation can be seen below, where K_x is the equilibrium constant of a reaction with substituent x, K_H is the equilibrium constant of a reaction when substituent x is a hydrogen, σ_x is the substituent constant, and ρ is the reaction constant (Equation 1).

$$\log K_x = \rho \sigma_x + \log K_H \quad [\text{Eq. 1}]$$

Hansch and his colleagues later published other works utilizing a successful computational approach to quantitatively model effects of substituents on biological activity.^{73, 74} Later, Hansch developed a new equation relating logP (log of octanol/water partition coefficient) and biological activity; this equation integrates the substituent constants from the Hammett equation (the σ values) along with constants obtained via regression analysis performed on an early version of a computer.⁷⁵ Predicting logP allowed researchers to theorize if a compound could pass through the cell membranes (the activity) by applying the equation to compounds not prepared yet in the lab.⁷⁶ Fitting this equation using a computer changed the idea that performing this math on a calculator was the only way to accomplish these predictions and paved the way for other scientists to expand their datasets by decreasing the time it would take to physically calculate the numbers by hand.⁷⁷ This groundbreaking new equation can be seen in Equation 2, where C is concentration, P is octanol/water partition coefficient, σ is the substituent constant, and k_1 , k_2 , k_3 , and k_4 are reaction rates.

$$\log(1/C) = -k_1(\log P)^2 + k_2(\log P) + k_3\sigma + k_4 \quad [\text{Eq. 2}]$$

These works showed scientists at the time that using a computer to accomplish these tasks was possible and paved the way for the field. Over the past fifty-seven years, the field of QSAR has expanded to become one of the most common approaches for correlating the chemical and physical properties of chemicals to the biological properties so essential to the field of medicinal chemistry.

A QSAR model is a function developed that relates some activity to molecular descriptors that reflect physical and chemical properties. Typically, such models have the general formula $Y = f(X)$, where Y is the activity being predicted and X are the descriptors.⁷⁸ Molecular descriptors commonly used in modern QSAR models are usually calculated using chemical structures as input and represent that structure at various levels.⁷⁹ Two-dimensional (2D) descriptors are derived from the structural formula and are conformationally independent—these descriptors are sometimes described as topological descriptors. Three-dimensional (3D) descriptors are derived from molecular geometry and are conformationally and stereochemically dependent—these descriptors are sometimes described as geometric descriptors. Some geometric descriptors are additionally dependent on molecular orientation within the coordinate space.

A common method of linear regression used in constructing QSAR models is known as partial least squares (PLS), which is designed so that the variables explain both the variation in the independent variables, X , and in the dependent variable, Y .⁸⁰ If the data resulting from high-throughput screening (HTS) gives binary activity results (active/inactive), then a binary QSAR model can be made.⁸¹ In a binary QSAR model, the activity of training compounds is represented as a binary value (either an active (1) or inactive (0)). Models constructed using binary activity data predict the probability of membership in the active group.^{81, 82} Both PLS and binary methodology were used in this work.

Methods

Database Compilation

Due to the abundance of known experimental data for therapeutics, information on the ADMET properties of interest (HIA, Caco-2 permeability, and percent protein binding) was compiled into a database. This database serves as the source for the training and validation compounds used in the computational models described herein. Compounds with published HIA data were obtained from a dataset compiled by Hou *et al.*⁸³ Caco-2 permeability data was obtained through two sources: DrugBank⁸⁴ and a dataset created by O'Hagan and Kell.⁸⁵ Percent Protein Binding was obtained through two sources: DrugBank⁸⁴ and TOXNET Hazardous Substances Data Bank.⁸⁶ A total of 562 compounds with known ADMET properties of interest were compiled into the database, but only those with reported values of all the properties of interest (HIA, Caco-2 permeability, and % protein binding) were utilized in model development. For cases where Caco-2 permeability had multiple values reported from different sources, the data was averaged, and standard deviations were calculated. For cases where percent protein binding was listed as a range of percentages, the median value was calculated and used.

Activity Categorization

Experimental data were categorized based on reported activity for each compound, i.e. a compound that did not move across the intestinal barrier readily was considered poorly or moderately absorbed, or a compound that is bound abundantly to plasma albumin was not considered to be distributed through the body well as it is difficult to dissociate from the protein. For HIA and Caco-2 permeability, compound activity is separated into three absorption categories. The *in vivo* absorption categories used herein are based on values outlined by

Radchenko *et al.*, where compounds with HIA greater than or equal to 70% are classified as well absorbed, between 70% and 30% classified as moderately absorbed, and less than or equal to 30% classified as poorly absorbed.⁸⁷ The Caco-2 permeability absorption cutoffs outlined by Yee in 1997 were used herein, where compounds with P_{app} values less than 1×10^{-6} cm/sec (log value less than -6.00 or 0-20%) were categorized as poorly absorbed, those with P_{app} between $1-10 \times 10^{-6}$ cm/sec (log value between -6.00 & -5.00 or 20-70%) were categorized as moderately absorbed, and those with P_{app} greater than 10×10^{-6} cm/sec (log value greater than -5.00 or 70-100%) were categorized as well absorbed.⁸⁸ It was noted that the cutoffs outlined for Caco-2 permeability vary from the cutoffs for HIA by 10%, making the moderately absorbed range slightly larger for Caco-2 permeability. The percent protein binding cutoffs were based on information collected by Scheife in 1989, where low protein binding was defined as being less than or equal to 80%, the clinically relevant moderate protein binding range was 80-85%, and above 85% was considered high binding where the drug's distribution is negatively impacted.⁸⁹ For the purpose of our models, we used two categories for protein binding, low and high binding, with the cutoff at 85%. These two categories were selected, rather than three categories, due to an insufficient amount of compounds within the middle category in the compiled database, and therefore, in our training and test set selection procedures.

Training and Test Set Selection

Compounds with ultimately desirable features were selected, i.e. representatives in all three absorption categories (poor, moderate, and well) for HIA and Caco-2 permeability and both low and high percent binding categories for protein binding, with no strong bias for the well absorbed compounds. We wanted to directly compare HIA and Caco-2 permeability, so

compounds having reported values in both systems were selected first. From that subset, we determined how many entries also had percent protein binding data. When selecting compounds for the test set, approximately 10% of the total number of training set compounds representative of each absorption category is typically chosen,⁹⁰ thus 10% of compounds selected were used as the test set.

HIA and Caco-2 Permeability Training and Test Sets

Of the 562 compiled compounds, only 114 had both HIA and Caco-2 permeability reported values. When examined in terms of absorption categories for HIA, 5 of the 114 compounds were poorly absorbed, 17 of the 114 compounds were moderately absorbed, and 92 of the compounds were well absorbed. As such, the number of compounds selected for both the training and test set was limited by the number of poorly absorbed compounds (a total of 5). It is preferable within the QSAR/QSPR field to use approximately one tenth (or 10%) as many compounds in the test set as there are in the training set, specifically compounds that are representative both in structure and activity.⁹⁰ Keeping with this convention, and with the limitations of the number of poorly absorbed compounds, 2 compounds from each absorption category were selected randomly to comprise the test set, with a total of 6 compounds comprising test set 1. Removing 2 of the 17 moderately absorbed compounds for the test set left 15 compounds in the training set in this category. To help limit the severity of the bias toward well absorbed compounds in the training set and to help fit the 10% convention, only 42 compounds of the 92 well absorbed were selected randomly, making the total number of compounds in the training set 60. This left 3 poorly absorbed compounds, 15 moderately absorbed compounds, and 42 well absorbed compounds in training set 1 based on the HIA

cutoffs. Training set 1 for the models, along with their HIA, Caco-2 permeability, and percent protein binding data can be found in Table 1. Compounds composing test set 1 compounds can be seen in Table 2.

Table 1. Training set 1 compounds and corresponding HIA, Caco-2 permeability, and protein binding data. Some compounds lacked protein binding data and are herein designated as No P.B. Some Caco-2 permeability values are presented as means and standard deviations of multiple reported values. Absorption categories and binding categories listed are based on each set of cutoffs as described in the Activity Categorization section of this chapter. Structures are shown in Table 7.

| Compound | % HIA | HIA Absorption Category | Log Caco-2 Permeability (# averaged values) | Caco-2 Absorption Category | % Protein Binding | Binding Category |
|---------------------|-------|-------------------------|---|----------------------------|-------------------|------------------|
| Aztreonam | 1 | Poorly | -6.19 | Poorly | 56 | Low |
| Amphotericin B | 3 | Poorly | -6.92 | Poorly | 90 | High |
| Ceftriaxone | 1 | Poorly | -6.89 | Poorly | 95 | High |
| Nadolol | 32 | Moderately | -5.56 ± 0.41 (10) | Moderately | 30 | Low |
| Atenolol | 50 | Moderately | -5.79 ± 0.69 (21) | Moderately | 11 | Low |
| Metformin | 54 | Moderately | -5.10 ± 0.83 (3) | Moderately | 0 | Low |
| Valsartan | 55 | Moderately | -6.28 ± 0.13 (5) | Poorly | 96 | High |
| Dipyridamole | 58 | Moderately | -5.21 ± 0.23 (5) | Moderately | 99 | High |
| Ziprasidone | 60 | Moderately | -4.91 | Well | 99 | High |
| Furosemide | 61 | Moderately | -6.06 ± 0.52 (15) | Poorly | 95 | High |
| Sulfasalazine | 62 | Moderately | -6.21 ± 0.4 (13) | Poorly | 99 | High |
| Sumatriptan | 67 | Moderately | -5.68 ± 0.14 (4) | Moderately | 18 | Low |
| Mibefradil | 69 | Moderately | -4.87 | Well | 99 | High |
| Erythromycin | 35 | Moderately | -6.03 ± 0.46 (10) | Poorly | 85 | Low |
| Metolazone | 64 | Moderately | -5.46 ± 0.28 (8) | Moderately | 33 | Low |
| Terbutaline | 63 | Moderately | -5.73 ± 0.6 (8) | Moderately | No P.B. | - |
| Sulpiride | 40 | Moderately | -6.29 ± 0.12 (3) | Poorly | No P.B. | - |
| Hydrochlorothiazide | 68 | Moderately | -5.85 ± 0.43 (6) | Moderately | 63 | Low |
| Ketoconazole | 75 | Well | -4.84 ± 0.14 (6) | Well | 99 | High |
| Acetaminophen | 85 | Well | -4.52 ± 0.46 (3) | Well | 25 | Low |
| Betaxolol | 90 | Well | -4.81 | Well | 50 | Low |
| Morphine | 90 | Well | -5.00 | Moderately | 35 | Low |
| Hydrocortisone | 91 | Well | -4.66 | Well | 95 | High |

Table 1 Continued. Training set 1 compounds and corresponding HIA, Caco-2 permeability, and protein binding data.

| Compound | % HIA | HIA Absorption Category | Log Caco-2 Permeability (# averaged values) | Caco-2 Absorption Category | % Protein Binding | Binding Category |
|-------------------|--------------|--------------------------------|--|-----------------------------------|--------------------------|-------------------------|
| Felodipine | 94 | Well | -4.97 ± 0.47 (2) | Well | 99 | High |
| Clonidine | 95 | Well | -4.56 ± 0.09 (4) | Well | 30 | Low |
| Metoprolol | 96 | Well | -4.51 ± 0.33 (23) | Well | 12 | Low |
| Ibuprofen | 98 | Well | -4.32 ± 0.07 (3) | Well | 95 | High |
| Caffeine | 99 | Well | -4.40 ± 0.08 (8) | Well | 31 | Low |
| Fluvastatin | 100 | Well | -5.31 ± 0.64 (8) | Moderately | 98 | High |
| Nicotine | 100 | Well | -4.65 ± 0.09 (2) | Well | 5 | Low |
| Tamoxifen | 100 | Well | -5.35 ± 0.92 (2) | Moderately | 99 | High |
| Testosterone | 100 | Well | -4.34 | Well | 98 | High |
| Alprenolol | 93 | Well | -4.58 ± 0.27 (5) | Well | 85 | Low |
| Ethinyl estradiol | 100 | Well | -3.42 | Well | 97 | High |
| Nitrendipine | 88 | Well | -4.77 | Well | 99 | High |
| Progesterone | 96 | Well | -4.37 | Well | 98 | High |
| Acebutolol | 85 | Well | -5.67 ± 0.27 (9) | Moderately | 26 | Low |
| Chlorpromazine | 98 | Well | -4.51 ± 0.32 (3) | Well | 94 | High |
| Bromazepam | 84 | Well | -4.40 | Well | 70 | Low |
| Bupropion | 87 | Well | -4.09 | Well | 84 | Low |
| Trovafloxacin | 88 | Well | -4.52 | Well | 76 | Low |
| Amphetamine | 90 | Well | -4.39 | Well | 28 | Low |
| Meloxicam | 90 | Well | -4.71 | Well | 99 | High |
| Phenytoln | 90 | Well | -4.49 ± 0.07 (4) | Well | 90 | High |
| Sulindac | 90 | Well | -5.29 ± 0.44 (6) | Moderately | 93 | High |
| Naloxone | 91 | Well | -4.62 ± 0.07 (3) | Well | No P.B. | - |
| Amrinone | 93 | Well | -5.06 ± 0.11 (5) | Moderately | 30 | Low |

Table 1 Continued. Training set 1 compounds and corresponding HIA, Caco-2 permeability, and protein binding data.

| Compound | % HIA | HIA Absorption Category | Log ₁₀ Caco-2 Permeability (# averaged values) | Caco-2 Absorption Category | % Protein Binding | Binding Category |
|--------------|-------|-------------------------|---|----------------------------|-------------------|------------------|
| Atropine | 94 | Well | -4.71 | Well | 18 | Low |
| Clozapine | 94 | Well | -4.80 | Well | 97 | High |
| Haloperidol | 100 | Well | -5.09 ± 0.42 (3) | Moderately | 92 | High |
| Hydralazine | 100 | Well | -4.84 ± 0.02 (3) | Well | 87 | High |
| Indomethacin | 100 | Well | -4.39 ± 0.38 (3) | Well | 98 | High |
| Ondansetron | 100 | Well | -3.96 | Well | 73 | Low |
| Diclofenac | 97 | Well | -4.27 | Well | 99 | High |
| Trimethoprim | 97 | Well | -4.42 ± 0.41 (5) | Well | 55 | Low |
| Naproxen | 99 | Well | -4.31 ± 0.18 (9) | Well | 98 | High |
| Omeprazole | 80 | Well | -4.30 | Well | 96 | High |
| Terbinafine | 80 | Well | -5.74 | Moderately | 90 | High |
| Domperidone | 95 | Well | -5.27 ± 0.02 (5) | Moderately | 92 | High |
| Flumazenil | 95 | Well | -4.53 | Well | 99 | High |

Table 2. Test set 1 compounds and the corresponding HIA, Caco-2 permeability, and percent protein binding data. Some compounds lacked protein binding data and are herein designated as No P.B. Some Caco-2 permeability values are presented as means and standard deviations of multiple reported values. Absorption categories and binding categories listed are based on each set of cutoffs as described in the Activity Categorization section of this chapter. Structures are shown in Table 7.

| Compound | % HIA | HIA Absorption Category | Log Caco-2 Permeability (# averaged values) | Caco-2 Absorption Category | % Protein Binding | Binding Category |
|-----------------|-------|-------------------------|---|----------------------------|-------------------|------------------|
| Mitoxantrone | 5 | Poorly | -5.04 ± 0.64 (3) | Moderately | 78 | Low |
| Lactulose | 1 | Poorly | -6.52 | Poorly | No P.B. | - |
| Fenoterol | 60 | Moderately | -6.11 | Poorly | No P.B. | - |
| Tranexamic acid | 55 | Moderately | -6.04 | Poorly | 3 | Low |
| Timolol | 95 | Well | -4.92 ± 0.31 (6) | Well | 10 | Low |
| Guanabenz | 78 | Well | -4.66 ± 0.44 (5) | Well | 90 | High |

A preference to have representatives of the test set well-spaced from boundaries between the HIA and Caco-2 absorption categories led to selection of training set 2 and test set 2. The distribution for training set 2 and test set 2 within the boundaries for absorption categories is shown in the Appendix, Figures A1 & A2. Training set 2 compounds for HIA and Caco-2 permeability can be seen in Table 3, along with the standard deviation for Caco-2 permeability (due to some of these values being the mean value recorded within the database). Test set 2 compounds for HIA and Caco-2 permeability can be seen in Table 4. Actual values used in the Caco-2 permeability QSAR models were the means, since multiple values for each compound could not be used. With the final rearrangement, 9 compounds were categorized as poorly absorbed with the Caco-2 permeability cutoffs, 16 compounds were moderately absorbed, and 35 compounds were defined as well absorbed.

Table 3. Training set 2 and the corresponding HIA and Caco-2 permeability data. Some Caco-2 permeability values are presented as means and standard deviations of multiple reported values. Absorption categories listed based on each set of cutoffs as described in the Activity Categorization section of this chapter. Structures are shown in Table 7.

| Compound | % HIA | HIA Absorption Category | Log Caco-2 Permeability (# averaged values) | Caco-2 Permeability Absorption Category |
|---------------------|-------|-------------------------|---|---|
| Aztreonam | 1 | Poorly | -6.19 | Poorly |
| Mitoxantrone | 5 | Poorly | -5.04 ± 0.64 (3) | Moderately |
| Ceftriaxone | 1 | Poorly | -6.89 | Poorly |
| Tranexamic Acid | 55 | Moderately | -6.04 | Poorly |
| Atenolol | 50 | Moderately | -5.79 ± 0.69 (21) | Moderately |
| Metformin | 54 | Moderately | -5.10 ± 0.83 (3) | Moderately |
| Valsartan | 55 | Moderately | -6.28 ± 0.13 (5) | Poorly |
| Dipyridamole | 58 | Moderately | -5.21 ± 0.23 (5) | Moderately |
| Ziprasidone | 60 | Moderately | -4.91 | Well |
| Furosemide | 61 | Moderately | -6.06 ± 0.52 (15) | Poorly |
| Sulfasalazine | 62 | Moderately | -6.21 ± 0.4 (13) | Poorly |
| Sumatriptan | 67 | Moderately | -5.68 ± 0.14 (4) | Moderately |
| Mibefradil | 69 | Moderately | -4.87 | Well |
| Erythromycin | 35 | Moderately | -6.03 ± 0.46 (10) | Poorly |
| Fenoterol | 60 | Moderately | -6.11 | Poorly |
| Terbutaline | 63 | Moderately | -5.73 ± 0.6 (8) | Moderately |
| Sulpiride | 40 | Moderately | -6.29 ± 0.12 (3) | Poorly |
| Hydrochlorothiazide | 68 | Moderately | -5.85 ± 0.43 (6) | Moderately |
| Ketoconazole | 75 | Well | -4.84 ± 0.14 (6) | Well |
| Acetaminophen | 85 | Well | -4.52 ± 0.46 (3) | Well |
| Betaxolol | 90 | Well | -4.81 | Well |
| Morphine | 90 | Well | -5.00 | Moderately |
| Hydrocortisone | 91 | Well | -4.66 | Well |
| Felodipine | 94 | Well | -4.97 ± 0.47 (2) | Well |
| Clonidine | 95 | Well | -4.56 ± 0.09 (4) | Well |
| Metoprolol | 96 | Well | -4.51 ± 0.33 (23) | Well |
| Ibuprofen | 98 | Well | -4.32 ± 0.07 (3) | Well |
| Caffeine | 99 | Well | -4.40 ± 0.08 (8) | Well |
| Fluvastatin | 100 | Well | -5.31 ± 0.64 (8) | Moderately |
| Nicotine | 100 | Well | -4.65 ± 0.09 (2) | Well |
| Tamoxifen | 100 | Well | -5.35 ± 0.92 (2) | Moderately |
| Testosterone | 100 | Well | -4.34 | Well |
| Alprenolol | 93 | Well | -4.58 ± 0.27 (5) | Well |
| Ethinyl Estradiol | 100 | Well | -3.42 | Well |
| Nitrendipine | 88 | Well | -4.77 | Well |

Table 3 Continued. Training set 2 and the corresponding HIA and Caco-2 permeability data.

| Compound | % HIA | HIA Absorption Category | Log Caco-2 Permeability (# averaged values) | Caco-2 Permeability Absorption Category |
|-----------------|--------------|--------------------------------|--|--|
| Progesterone | 96 | Well | -4.37 | Well |
| Acebutolol | 85 | Well | -5.67 ± 0.27 (9) | Moderately |
| Chlorpromazine | 98 | Well | -4.51 ± 0.32 (3) | Well |
| Bromazepam | 84 | Well | -4.4 | Well |
| Timolol | 95 | Well | -4.92 ± 0.31 (6) | Well |
| Trovafloxacin | 88 | Well | -4.52 | Well |
| Amphetamine | 90 | Well | -4.39 | Well |
| Meloxicam | 90 | Well | -4.71 | Well |
| Phenytoin | 90 | Well | -4.49 ± 0.07 (4) | Well |
| Sulindac | 90 | Well | -5.29 ± 0.44 (6) | Moderately |
| Naloxone | 91 | Well | -4.62 ± 0.07 (3) | Well |
| Amrinone | 93 | Well | -5.06 ± 0.11 (5) | Moderately |
| Atropine | 94 | Well | -4.71 | Well |
| Clozapine | 94 | Well | -4.80 | Well |
| Haloperidol | 100 | Well | -5.09 ± 0.42 (3) | Moderately |
| Hydralazine | 100 | Well | -4.84 ± 0.02 (3) | Well |
| Indomethacin | 100 | Well | -4.39 ± 0.38 (3) | Well |
| Ondansetron | 100 | Well | -3.96 | Well |
| Diclofenac | 97 | Well | -4.27 | Well |
| Trimethoprim | 97 | Well | -4.42 ± 0.41 (5) | Well |
| Naproxen | 99 | Well | -4.31 ± 0.18 (9) | Well |
| Omeprazole | 80 | Well | -4.30 | Well |
| Terbinafine | 80 | Well | -5.74 | Moderately |
| Domperidone | 95 | Well | -5.27 ± 0.02 (5) | Moderately |
| Flumazenil | 95 | Well | -4.53 | Well |

Table 4. Test set 2 compounds and the corresponding HIA and Caco-2 permeability data. Some Caco-2 permeability values are presented as means and standard deviations of multiple reported values. Absorption categories listed based on each set of cutoffs as described in the Activity Categorization section of this chapter. Structures are shown in Table 7.

| Compound | % HIA | HIA Absorption Category | Log Caco-2 Permeability (# averaged values) | Caco-2 Permeability Absorption Category |
|----------------|-------|-------------------------|---|---|
| Lactulose | 1 | Poorly | -6.52 | Poorly |
| Amphotericin B | 3 | Poorly | -6.92 | Poorly |
| Metolazone | 64 | Moderately | -5.46 ± 0.28 (8) | Moderately |
| Nadolol | 32 | Moderately | -5.56 ± 0.41 (10) | Moderately |
| Guanabenz | 78 | Well | -4.66 ± 0.44 (5) | Well |
| Bupropion | 87 | Well | -4.09 | Well |

Percent Protein Binding Training and Test Set

In terms of compiled percent protein binding data, 61 of the 66 compounds selected for training and test set 1 had reported values, with 57 in the training set and 4 in the test set (training and test set 1 shown in Tables 1 & 2). Training set 1 contained 25 low binding compounds and 32 high binding compounds. Test set 1 had 3 low binding compounds and 1 high binding compound. A few compounds in the test set were exchanged with training set compounds to better represent the training set, making training set 3 and test set 3 for protein binding, which differ in compounds from training set and test set 2 for HIA and Caco-2 permeability. Training set 3 had 27 low binding compounds and 30 high binding compounds. Test set 3 comprised 1 low binding compound and 3 high binding compounds. A final list of compounds used in the training and test sets 3 for protein binding QSAR models can be seen in Tables 5 and 6 respectively. Structures for all compounds within training and test sets 1-3 are shown in Table 7.

Table 5. Training set 3 compounds selected for % protein binding QSAR models. Protein binding category based on cutoffs are also shown as described in the Activity Categorization section of this chapter. Structures are shown in Table 7.

| Compound | % Protein Binding | Binding Category |
|---------------------|--------------------------|-------------------------|
| Metformin | 0 | Low |
| Tranexamic Acid | 3 | Low |
| Nicotine | 5 | Low |
| Timolol | 10 | Low |
| Atenolol | 11 | Low |
| Metoprolol | 12 | Low |
| Sumatriptan | 18 | Low |
| Atropine | 18 | Low |
| Acetaminophen | 25 | Low |
| Acebutolol | 26 | Low |
| Amphetamine | 28 | Low |
| Nadolol | 30 | Low |
| Clonidine | 30 | Low |
| Amrinone | 30 | Low |
| Caffeine | 31 | Low |
| Metolazone | 33 | Low |
| Morphine | 35 | Low |
| Betaxolol | 50 | Low |
| Trimethoprim | 55 | Low |
| Aztreonam | 56 | Low |
| Hydrochlorothiazide | 63 | Low |
| Bromazepam | 70 | Low |
| Ondansetron | 73 | Low |
| Trovafloxacin | 76 | Low |
| Bupropion | 84 | Low |
| Erythromycin | 85 | Low |
| Alprenolol | 85 | Low |
| Amphotericin B | 90 | High |
| Phenytoin | 90 | High |
| Terbinafine | 90 | High |
| Haloperidol | 92 | High |
| Domperidone | 92 | High |
| Chlorpromazine | 94 | High |
| Ceftriaxone | 95 | High |
| Furosemide | 95 | High |
| Hydrocortisone | 95 | High |
| Ibuprofen | 95 | High |
| Valsartan | 96 | High |

Table 5 Continued. Training set 3 compounds selected for % protein binding QSAR models.

| Compound | % Protein Binding | Binding Category |
|-------------------|--------------------------|-------------------------|
| Omeprazole | 96 | High |
| Ethinyl Estradiol | 97 | High |
| Clozapine | 97 | High |
| Fluvastatin | 98 | High |
| Testosterone | 98 | High |
| Progesterone | 98 | High |
| Indomethacin | 98 | High |
| Naproxen | 98 | High |
| Dipyridamole | 99 | High |
| Ziprasidone | 99 | High |
| Sulfasalazine | 99 | High |
| Mibefradil | 99 | High |
| Ketoconazole | 99 | High |
| Felodipine | 99 | High |
| Tamoxifen | 99 | High |
| Nitrendipine | 99 | High |
| Meloxicam | 99 | High |
| Diclofenac | 99 | High |
| Flumazenil | 99 | High |

Table 6. Test set 3 compounds selected for % protein binding QSAR models. Protein binding category based on cutoffs are also shown as described in the Activity Categorization section of this chapter. Structures are shown in Table 7.

| Compound | % Protein Binding | Binding Category |
|-----------------|--------------------------|-------------------------|
| Mitoxantrone | 78 | Low |
| Hydralazine | 87 | High |
| Guanabenz | 90 | High |
| Sulindac | 93 | High |

Table 7. Structures of all compounds selected to build QSAR models. All structures are shown ionized at physiological pH 7.4. * symbol indicates set 1, ** symbol indicates set 2, * symbol indicates set 3, highlighted symbol means the compound was in the test set for that dataset. A + symbol indicates compounds for which stereochemistry was used in the modeling process or model assessment, but no stereochemistry was reported to accompany the experimental data.**

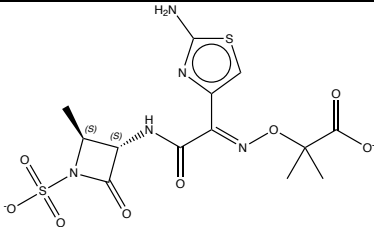
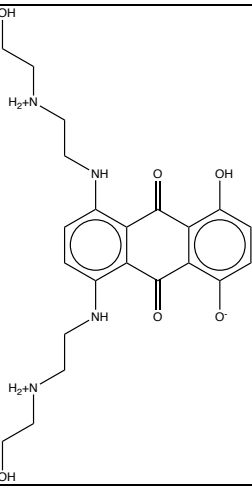
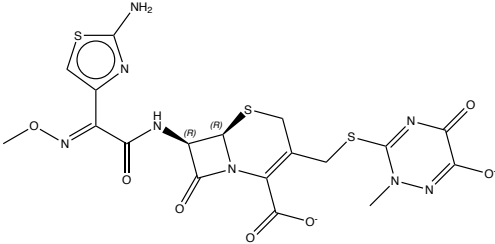
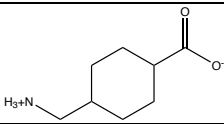
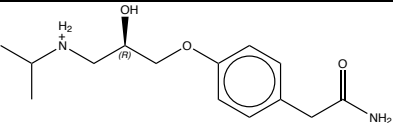
| # | Compound | Structure (pH 7.4) |
|---|--------------------------------|--|
| 1 | Aztreonam (*/**/****) |  |
| 2 | Mitoxantrone (*/**/****) |  |
| 3 | Ceftriaxone (*/**/****) |  |
| 4 | Tranexamic Acid (*/**/****) |  |
| 5 | Atenolol+ (*/**/****) |  |

Table 7 Continued. Structures of all compounds selected to build QSAR models.

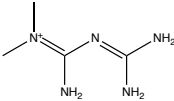
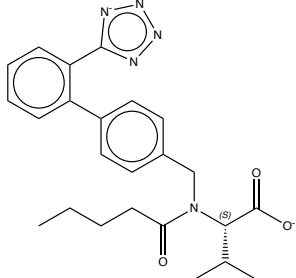
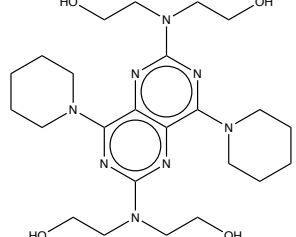
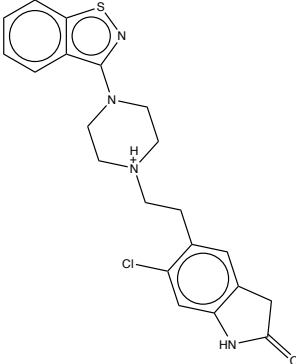
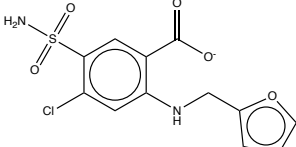
| # | Compound | Structure (pH 7.4) |
|----|----------------------------|--|
| 6 | Metformin (*/**/***) |  |
| 7 | Valsartan (*/**/***) |  |
| 8 | Dipyridamole (*/**/***) |  |
| 9 | Ziprasidone (*/**/***) |  |
| 10 | Furosemide (*/**/***) |  |

Table 7 Continued. Structures of all compounds selected to build QSAR models.

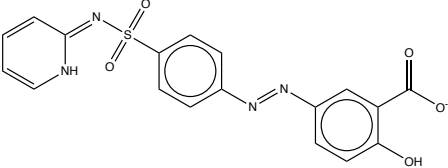
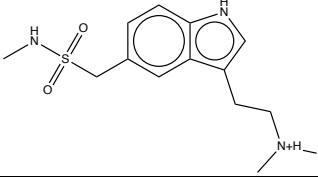
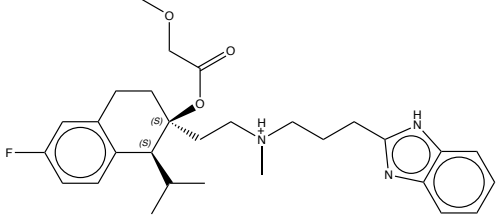
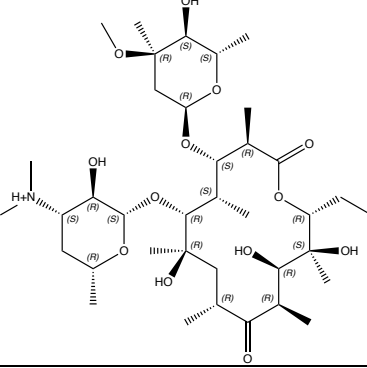
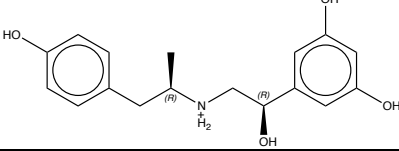
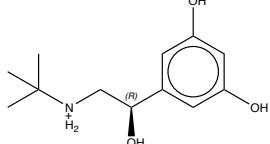
| # | Compound | Structure (pH 7.4) |
|----|-----------------------------|--|
| 11 | Sulfasalazine (*/**/***) |  |
| 12 | Sumatriptan (*/**/***) |  |
| 13 | Mibefradil (*/**/***) |  |
| 14 | Erythromycin (*/**/***) |  |
| 15 | Fenoterol+ (*/**) |  |
| 16 | Terbutaline+ (*/**) |  |

Table 7 Continued. Structures of all compounds selected to build QSAR models.

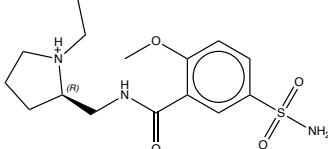
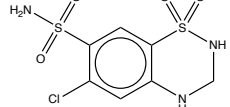
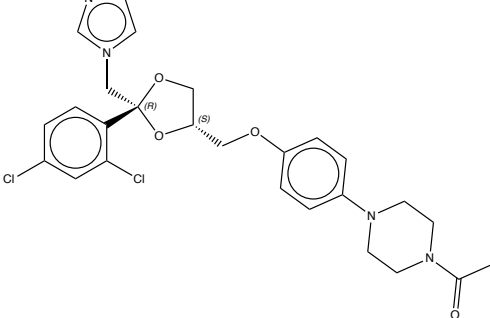
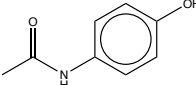
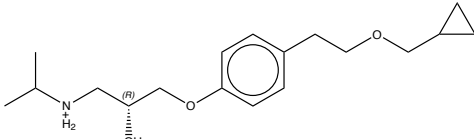
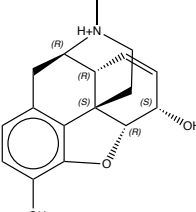
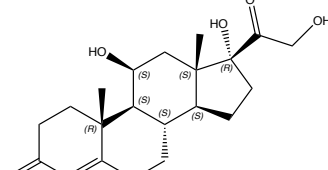
| # | Compound | Structure (pH 7.4) |
|----|------------------------------------|--|
| 17 | Sulpiride+ (*/**) |  |
| 18 | Hydrochlorothiazide (*/**/****) |  |
| 19 | Ketoconazole (*/**/****) |  |
| 20 | Acetaminophen (*/**/****) |  |
| 21 | Betaxolol+ (*/**/****) |  |
| 22 | Morphine (*/**/****) |  |
| 23 | Hydrocortisone (*/**/****) |  |

Table 7 Continued. Structures of all compounds selected to build QSAR models.

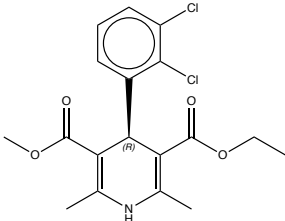
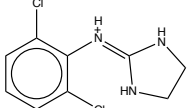
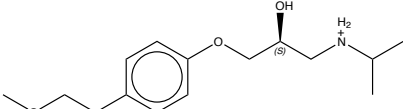
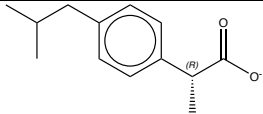
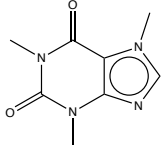
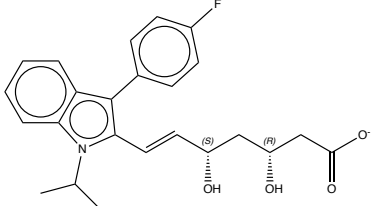
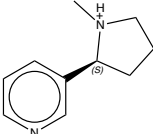
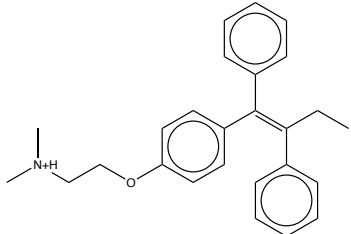
| # | Compound | Structure (pH 7.4) |
|----|---------------------------|--|
| 24 | Felodipine+ (*/**/***) |  |
| 25 | Clonidine (*/**/***) |  |
| 26 | Metoprolol+ (*/**/***) |  |
| 27 | Ibuprofen+ (*/**/***) |  |
| 28 | Caffeine (*/**/***) |  |
| 29 | Fluvastatin (*/**/***) |  |
| 30 | Nicotine (*/**/***) |  |
| 31 | Tamoxifen (*/**/***) |  |

Table 7 Continued. Structures of all compounds selected to build QSAR models.

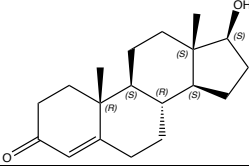
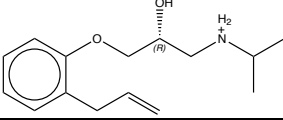
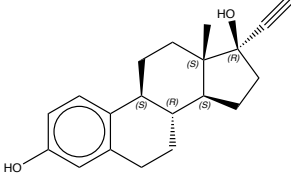
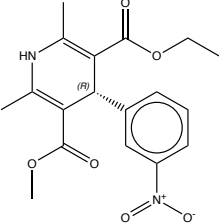
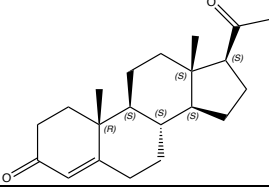
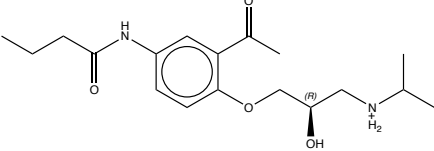
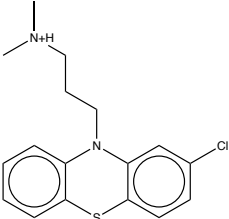
| # | Compound | Structure (pH 7.4) |
|----|---------------------------------|--|
| 32 | Testosterone (*/**/***) |  |
| 33 | Alprenolol (*/**/***) |  |
| 34 | Ethinyl Estradiol (*/**/***) |  |
| 35 | Nitrendipine+ (*/**/***) |  |
| 36 | Progesterone (*/**/***) |  |
| 37 | Acebutolol+ (*/**/***) |  |
| 38 | Chlorpromazine (*/**/***) |  |

Table 7 Continued. Structures of all compounds selected to build QSAR models.

| # | Compound | Structure (pH 7.4) |
|----|---------------------------|--------------------|
| 46 | Naloxone (*/**) | |
| 47 | Amrinone (*/**/**) | |
| 48 | Atropine+ (*/**/**) | |
| 49 | Clozapine (*/**/**) | |
| 50 | Haloperidol (*/**/**) | |
| 51 | Hydralazine+ (*/**/**) | |

Table 7 Continued. Structures of all compounds selected to build QSAR models.

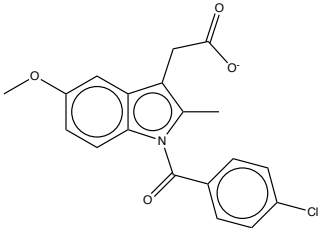
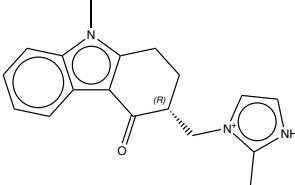
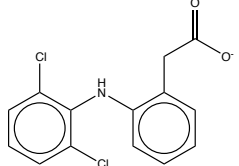
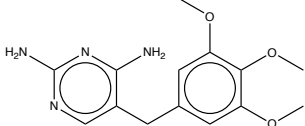
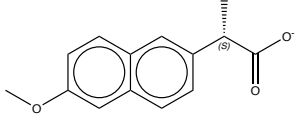
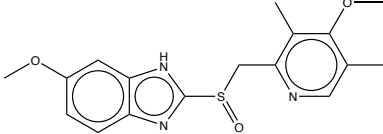
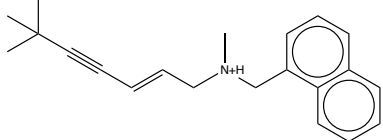
| # | Compound | Structure (pH 7.4) |
|----|---|--|
| 52 | Indomethacin (*/**/****) |  |
| 53 | Ondansetron ⁺ (*/**/****) |  |
| 54 | Diclofenac (*/**/****) |  |
| 55 | Trimethoprim (*/**/****) |  |
| 56 | Naproxen (*/**/****) |  |
| 57 | Omeprazole (*/**/****) |  |
| 58 | Terbinafine (*/**/****) |  |

Table 7 Continued. Structures of all compounds selected to build QSAR models.

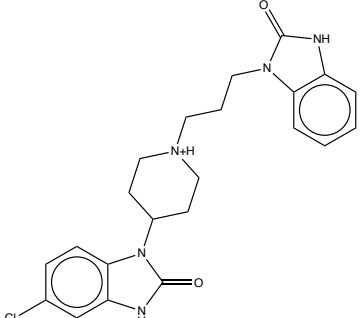
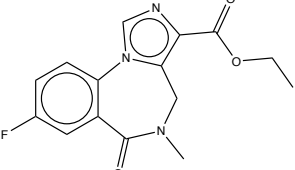
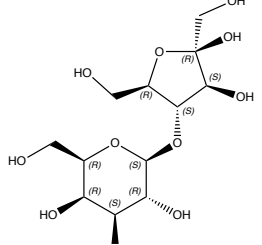
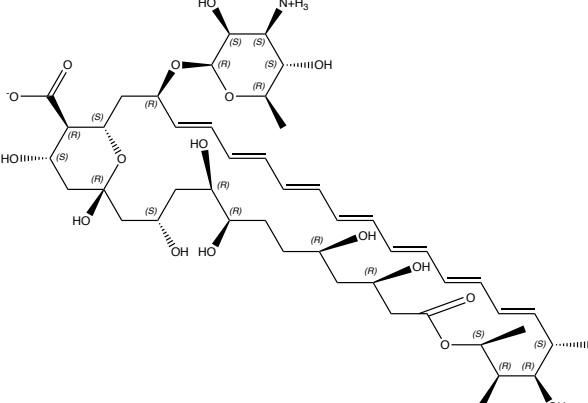
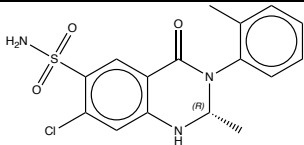
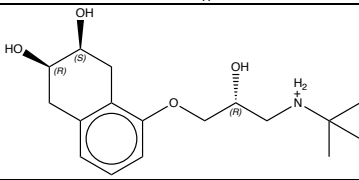
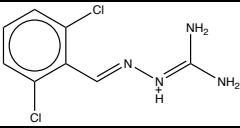
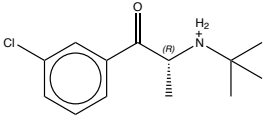
| # | Compound | Structure (pH 7.4) |
|----|------------------------------|--|
| 59 | Domperidone (*/**/***) |  |
| 60 | Flumazenil (*/**/***) |  |
| 61 | Lactulose (*/**) |  |
| 62 | Amphotericin B (*/**/***) |  |

Table 7 Continued. Structures of all compounds selected to build QSAR models.

| # | Compound | Structure (pH 7.4) |
|----|----------------------------|---|
| 63 | Metolazone+ (*/**/****) |  |
| 64 | Nadolol+ (*/**/****) |  |
| 65 | Guanabenz (*/**/****) |  |
| 66 | Bupropion+ (*/**/****) |  |

Chemical Diversity Measures

The diversity of the dataset of compounds selected for the training and test sets was determined to ensure that a broad chemical space was represented in an effort to ensure that the models were suitable for a variety of chemical structures. Diversity was assessed based on methods suggested by Gonzalez-Medina *et al.*⁹¹ Molecular ACCess System (MACCS) keys provided a numerical fingerprint of the structures, and a Tanimoto coefficient comparison between sets of MACCS keys was computed using the MOE 2016 software package.⁹² Each MACCS key contains a series of 166 bits, used to indicate the presence (bit value = 1) or absence (bit value = 0) of small substructures of 1-10 non-hydrogen atoms.⁹³ The Tanimoto coefficient is a similarity measurement comparing two fingerprints, in this case MACCS keys, which varies

between a minimum of 0 and a maximum of 1, with higher values indicating higher similarity.⁹¹

⁹⁴ The formula for the Tanimoto coefficient can be seen in Equation 3, where S is similarity, a is the number of bits on molecule A, b is the number of bits on molecule B, and c is the number of bits on both molecules A & B.

$$S_{A/B} = \frac{c}{[a+b-c]} \quad [\text{Eq. 3}]$$

A similarity matrix of Tanimoto coefficients was calculated using the pH4_SimilarityMatrix function in the MOE SVL command window. Using this function, similarity matrices were calculated for the following compound sets: training and test sets 1-3 both internally and against each other, along with each absorption category and each binding category internally and against each other. From these similarity matrices, histograms were plotted and utilized to visualize dataset diversity.

QSAR Model Development, Internal Validation, and External Validation

Figure 5 shows the workflow for QSAR model development, validation, and application. This shows the overall process utilized in these methods, along with the overall goal of the models. The models were constructed using compounds with known ADMET properties. The constructed models were then applied to the test set compounds (also with known ADMET properties) to compare the predicted ADMET property to the actual ADMET property reported through the validation process. The goal was to develop validated models that can be used to make predictions on compounds of interest to help guide future ligand/inhibitor discovery efforts.

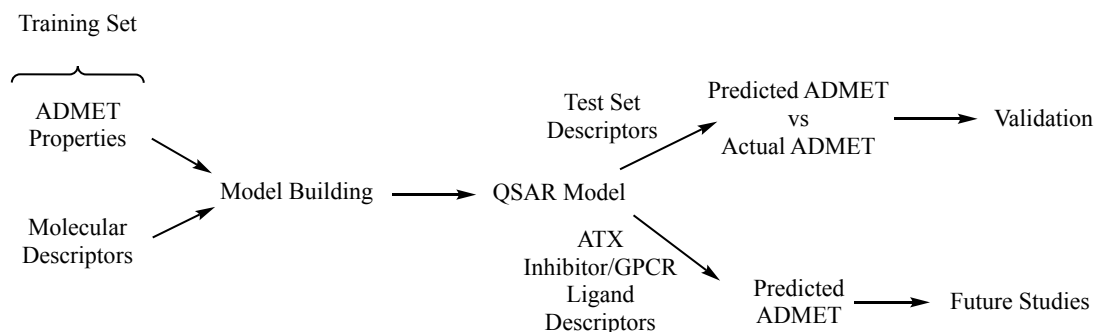


Figure 5. Workflow of the development, validation, and application of ADMET QSAR models. First, a training set was defined containing compounds with known ADMET properties, and molecular descriptors were selected to help build the model. The QSAR model was used to predict the ADMET properties for the training set using the molecular descriptors calculated and the predicted ADMET properties were compared to the known ADMET values of the training set to assess fit of the model to the training data. Second, the QSAR model was evaluated using the test set descriptors, where the predicted ADMET properties were compared to the measured ADMET properties of the test set to externally validate the QSAR model.

All compounds were modeled in the ionization state expected at physiological pH (7.4) using the wash function in MOE, followed by visual inspection/validation of all processed structures. There are different molecular descriptors relevant to this work, 2D descriptors and i3D descriptors. Two-dimensional (2D) molecular descriptors are calculated based on a compound's 2 dimensional structural formula, while i3D molecular descriptors are calculated from the 3 dimensional geometry of the molecule independent of position in the x,y,z coordinate system.⁷⁹ If 3D molecular descriptors were used in a model, they were calculated using the lowest energy conformation of the compounds, which were determined through a full conformational search. Conformational searches were performed using the Merck Molecular Force Field 94x (MMFF94x) as implemented in MOE on the University of Memphis High-Performance Computing (HPC) system using two solvation treatments, gas phase and the Born solvation model. The MMFF94x forcefield is a modification of the MMFF94s⁹⁵ forcefield with planar conjugated nitrogen atoms. The lowest energy conformation from each conformational

search was then used in each model that included descriptors dependent upon conformation (i3D and 2D/i3D combined). In this particular work 5 different models were made for each ADMET property examined using varying combinations of molecular descriptors (2D, i3D from two solvation models, and 2D/i3D combined from two solvation models) to determine which combination produced the most accurate predictions. All descriptors possible for each descriptor-type model were calculated (206 2D descriptors, 117 i3D descriptors, and a total of 323 2D/i3D descriptors), except semi-empirical i3D descriptors. Generally, only utilizing 1/5 as many descriptors as training set compounds is preferable to avoid overfitting the data when performing linear regression models, which is often referred to as the “Topliss-Costello rule”.^{66, 96} Following this trend, no more than 12 descriptors were used in each model described herein.

Contingency analysis was used to prune the molecular descriptors so only those most correlated to the activity field in question (HIA, Caco-2 permeability, or % protein binding), but those significantly correlated to each other were not used. Contingency analysis correlates the molecular descriptors to the activity field using a series of statistical measurements—contingency coefficient (C field), Cramer’s V (V field), entropic uncertainty (U field), and the correlation coefficient (R field).⁹⁷⁻¹⁰¹ For HIA and % protein binding, the descriptors were pruned by eliminating those showing values below 0.9 in the C field. For Caco-2 permeability, the descriptors were pruned by eliminating those giving values below 0.8 in the C field. To ensure the descriptors were not too correlated to each other, a correlation matrix was generated using MOE. Only descriptors with 80% correlation and below were kept. If two descriptors were more than 80% correlated, the descriptor least correlated to the other descriptors was kept.

QSAR models were generated based on the selected descriptors using the QuaSAR-Model mode in MOE. Linear models, using the partial least squares method, were made for both

HIA and Caco-2 permeability, while binary models were made for % protein binding. The model files were internally validated on the training set compounds through a cross-validation method called leave-one-out (LOO). LOO derives a model with the remaining data after each data value is left out in turn.^{76, 102} This means one compound is left out at a time and the remaining compounds were used as the training set to derive the model. The one compound left out was used as a single compound test set to validate the model. This process was repeated with all compounds in the training set to help validate the model by calculating the log of the sum of all the predicted values.¹⁰² The model files were then used to make predictions on the test set compounds through the Model Evaluate method in MOE. The model predictions were then compared to the experimental data based on the category corresponding to the predicted values to determine the accuracy of all 5 model types for HIA, Caco-2 permeability, and % protein binding.

Examination of Commonly Mispredicted Compounds

Within each model and for each activity (HIA, Caco-2 permeability, and % protein binding), mispredicted compounds were examined to see if any common features of poorly predicted compounds could be identified and used to define an appropriate scope of chemical structures for which the models were best suited. Similarities of mispredicted compounds were determined as described previously in the Chemical Diversity Measures section of this chapter.

Results and Discussion

Dataset Diversity

A histogram representing the diversity of all compounds in the compiled compound database is shown below in Figure 6. Histograms representing chemical diversity in individual training and test sets can be seen in the Appendix, Figures A3-A21. In all cases, the majority of similarities fall within the 0-50% range of similarity, thus all subsets represented diverse datasets.

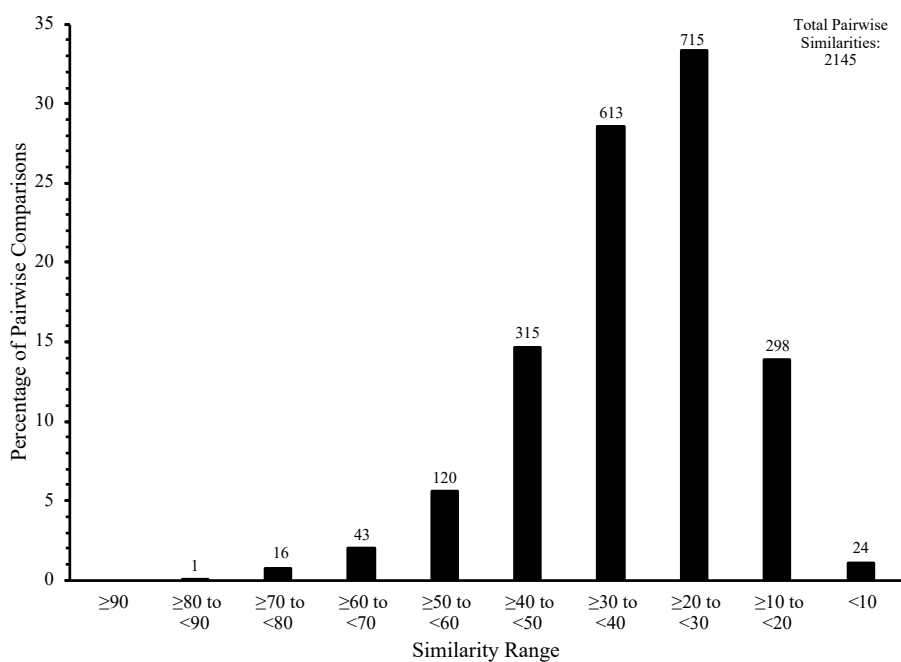


Figure 6. Representative histogram of all compound chemical diversity calculated from MACCS keys using Tanimoto coefficients. Number of pairwise comparisons within each 10% bin are on top of the bar along with the total of pairwise similarities to the right of the figure.

Comparison of 5 Descriptor Model Types

To determine the accuracy of the models constructed with different descriptor types used or solvation models, the number of compounds correctly predicted within each category (absorption category for HIA or Caco-2 permeability and high/low binding for protein binding) was calculated and plotted against the reported experimental value for each descriptor model type at each stage of modeling. Figure 7 shows the model comparison for human intestinal absorption at the model development stage and internal validation stage. Figure 8 shows the model comparison for human intestinal absorption at the external validation stage.

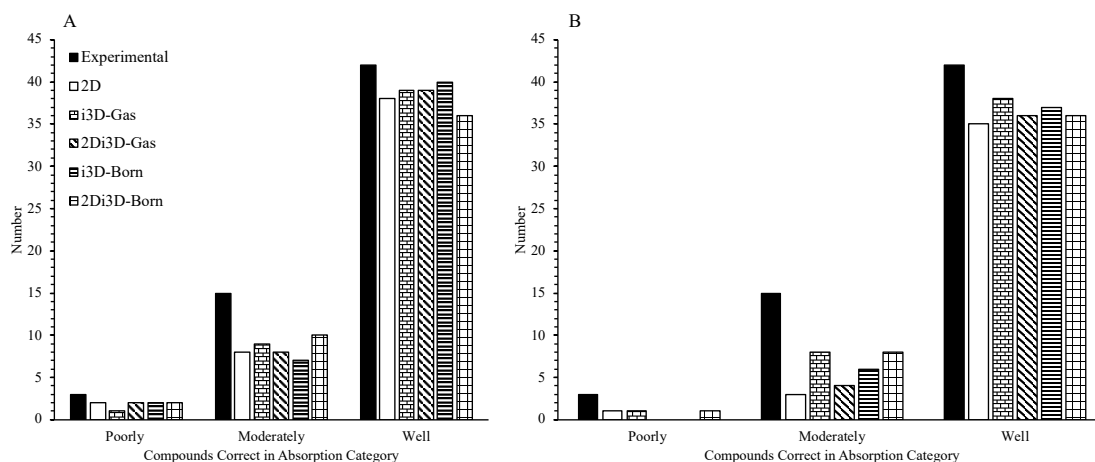


Figure 7. Model comparisons at the model development stage (panel A) and internal validation stage (panel B) for human intestinal absorption data. Experimental values are in black, 2D descriptor model values in open bars, i3D descriptor models in gas phase values in brick, 2D3D descriptor combination in gas phase in diagonal lines, i3D descriptor models in Born solvation values in horizontal lines, and 2D3D descriptor combination in Born solvation values in squares.

Based on the results shown in Figure 7 compounds in the well absorbed and poorly absorbed categories were more accurately predicted than those compounds in the moderately absorbed category during the model development stage. The well absorbed and poorly absorbed

compounds all closely match the experimental data (with at most only 6 of the 42 well absorbed compounds and 1 of 3 poorly absorbed compounds incorrectly categorized). Compounds in the poorly and moderately absorbed categories were poorly predicted in the internal validation stage, (at most all 3 poorly absorbed incorrectly categorized and 12 of 15 moderately absorbed were incorrectly categorized). Of the descriptor types used to create the models, the i3D Born solvated model performed slightly better than the other models for model development. In the internal validation stage, the i3D gas phase model performed slightly better than the other models.

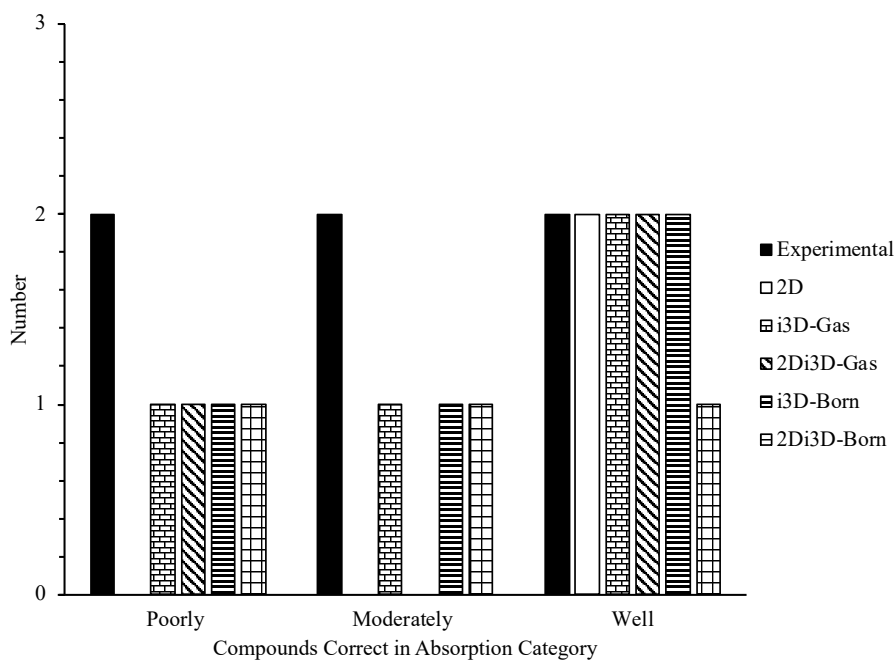


Figure 8. Model comparison at the external validation stage for human intestinal absorption data. Experimental values shown in black, 2D descriptor model values shown in open bars, i3D descriptor model values in the gas phase shown in bricks, 2D3D descriptor combination model values shown in diagonal bars, i3D descriptor model values in the Born solvation shown in horizontal lines, and 2D3D descriptor combination model values shown in squares.

The results shown in Figure 8 indicate compounds in the well absorbed category were most accurately predicted. Compounds in the moderately absorbed category were most poorly predicted, with two of five models failing to correctly predict either test compound in this absorption category. Of the descriptor types used to create the models, the i3D gas phase and i3D born solvated models performed equally.

Based on the results shown in all three comparisons for the models predicting HIA, there is little difference between each set of models. Figure 9 shows the model comparison for Caco-2 permeability at the model development stage and internal validation stage. Figure 10 shows the model comparison for Caco-2 permeability at the external validation stage.

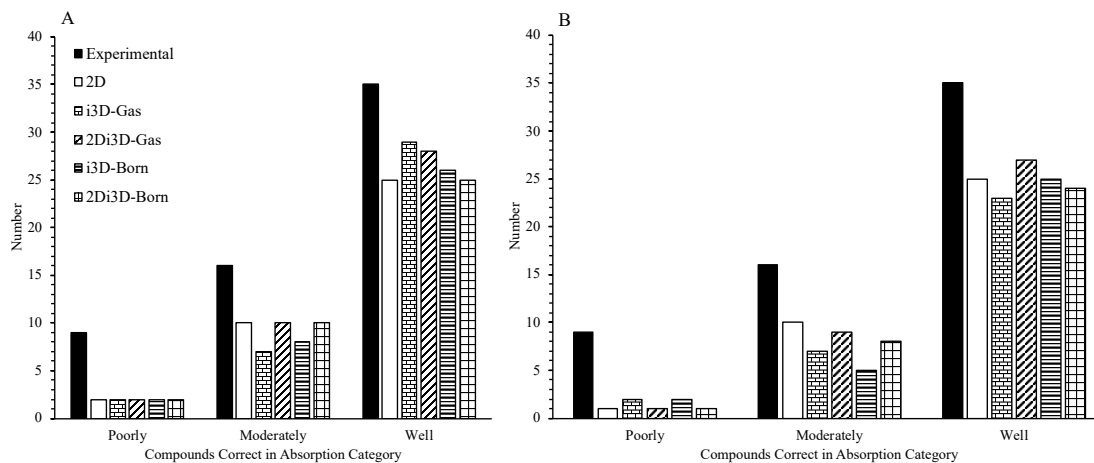


Figure 9. Model comparisons at the model development stage (panel A) and internal validation stage (panel B) for Caco-2 permeability data. Experimental values are in black, 2D descriptor model values in open bars, i3D descriptor models in gas phase values in brick, 2D3D descriptor combination in gas phase in diagonal lines, i3D descriptor models in Born solvation values in horizontal lines, and 2D3D descriptor combination in Born solvation values in squares.

Figure 9 indicates compounds in the well absorbed and moderately absorbed categories were more accurately predicted than those compounds in the poorly absorbed category during

the model development stage. Compounds in the poorly absorbed category performed the worst in both the model development and internal validation stage. Of the descriptor types used to create the models, the i3D-Born solvated model performed slightly better at the model development stage, but the 2D descriptor model performed slightly better at the internal validation stage.

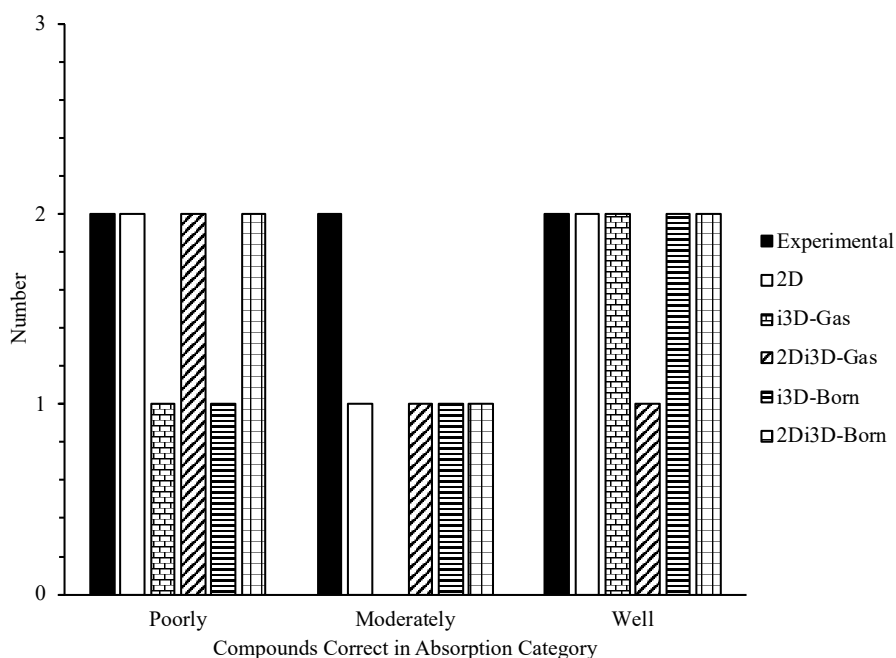


Figure 10. Model comparison at the external validation stage for Caco-2 permeability data. Experimental values shown in black bars, 2D descriptor model values shown in open bars, i3D descriptor model values for the gas phase are shown in brick bars, 2D3D descriptor combination values for gas phase shown in diagonal bars, i3D descriptor values for the Born solvation shown in horizontal bars, and 2D3D descriptor values for the Born solvation shown in square bars.

Based on the results shown in Figure 10 compounds in the well absorbed category were the most accurately predicted, with four out of five models correctly predicting all of the well-absorbed compounds. Compounds in the poorly absorbed category were similarly predicted, with three out of five models predicting all of the poorly absorbed compounds. Compounds in the

moderately absorbed category were predicted the worst, with no model correctly predicting both compounds and one model failing to correctly predict either compound in this category. This could be due to the large standard deviation some of the compounds had for Caco-2 permeability when they were means of multiple reported values.

The results shown in all three comparisons for the models predicting Caco-2 permeability, there is once again little difference between each set of models. Figure 11 shows the model comparison for protein binding at the model development stage and internal validation stage. Figure 12 shows the model comparison for protein binding at the external validation stage.

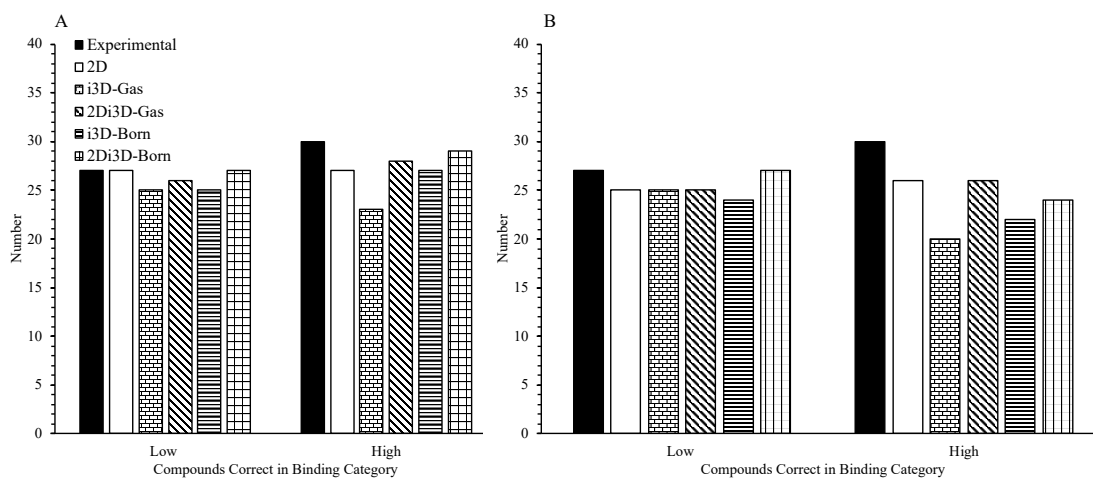


Figure 11. Model comparisons at the model development stage and internal validation stage for protein binding data. Experimental values shown in black bars, 2D descriptor model values shown in open bars, i3D descriptor model values in gas phase shown in brick bars, 2Di3D descriptor model values in gas phase shown in diagonal bars, i3D descriptor model values in Born solvation shown in horizontal line bars, and 2Di3D descriptor model values in Born solvation shown in square bars.

The results shown in Figure 11, indicate the majority of models performed similarly during both the model development stage and internal validation stage. Compounds in the high binding category performed slightly worse in the internal validation stage when compared to the

model development. Due to similar model performance similarly, it is difficult to select the best model to use based on these two stages of model development.

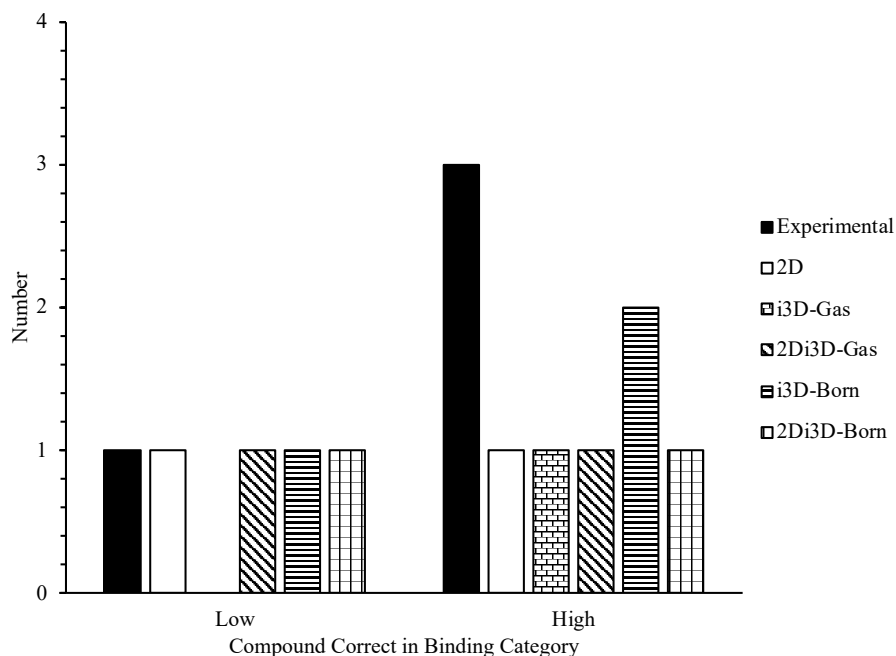


Figure 12. Model comparison at the external validation stage for protein binding data. Experimental values shown in black bars, 2D descriptor model values shown in open bars, i3D descriptor model values in gas phase shown in brick bars, 2D3D descriptor combination model values in gas phase shown in diagonal bars, i3D descriptor model values in Born solvation shown in horizontal line bars, and 2D3D descriptor combination model values in Born solvation shown in square bars.

Figure 12 illustrates model performance during the external validation stage. The model that performed the best overall at this stage was the i3D Born model, with 3 of 4 test compounds correctly categorized. However, the 2D/i3D models using both solvation conditions correctly categorized 2 out of 4 compounds. Based on the protein binding predictions results shown in all three comparisons, there is little difference between the models. Tables 8-10 give a more quantitative representation of the accuracy of each model comparison for HIA, Caco-2 permeability, and protein binding respectively.

Table 8. Comparison of all five models made with different descriptor types and solvation methods to predict HIA. Bold box indicates the model with the best predictions overall. Compounds in each absorption category in the training set (model development & internal validation): 3 poorly absorbed, 15 moderately absorbed, 42 well absorbed. Compounds in each absorption category in the test set (external validation): 2 poorly absorbed, 2 moderately absorbed, 2 well absorbed.

| | | Gas Phase | | | | | | Born Solvated | | | |
|---------------------|------------|-----------|-----------|---------|-----------|---------|-----------|---------------|-----------|---------|-----------|
| | | 2D | | i3D | | 2D/3D | | i3D | | 2D/3D | |
| | | Correct | Incorrect | Correct | Incorrect | Correct | Incorrect | Correct | Incorrect | Correct | Incorrect |
| Model Development | Poorly | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 |
| | Moderately | 8 | 7 | 9 | 6 | 8 | 7 | 7 | 8 | 10 | 5 |
| | Well | 38 | 4 | 39 | 3 | 39 | 3 | 40 | 2 | 36 | 6 |
| Internal Validation | Poorly | 1 | 2 | 1 | 2 | 0 | 3 | 0 | 3 | 1 | 2 |
| | Moderately | 3 | 12 | 8 | 7 | 4 | 11 | 6 | 9 | 8 | 7 |
| | Well | 35 | 7 | 38 | 4 | 36 | 6 | 37 | 5 | 36 | 6 |
| External Validation | Poorly | 0 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | Moderately | 0 | 2 | 1 | 1 | 0 | 2 | 1 | 1 | 1 | 1 |
| | Well | 2 | 0 | 2 | 0 | 2 | 0 | 2 | 0 | 1 | 1 |

Table 9. Comparison of all five models made with different descriptor types and solvation methods for Caco-2 permeability. Bold box indicates the model with the best predictions overall. Compounds in each absorption category in the training set (model development & internal validation): 9 poorly absorbed, 16 moderately absorbed, 35 well absorbed. Compounds in each absorption category in the test set (external validation): 2 poorly absorbed, 2 moderately absorbed, 2 well absorbed. * means 1 compound 2 absorption categories away. + means 1 compound previous category (for moderately absorbed compounds).

| | | Gas Phase | | | | | | Born Solvated | | | |
|---------------------|------------|-----------|-----------|---------|-----------|---------|-----------|---------------|-----------|---------|-----------|
| | | 2D | | i3D | | 2D/3D | | i3D | | 2D/3D | |
| | | Correct | Incorrect | Correct | Incorrect | Correct | Incorrect | Correct | Incorrect | Correct | Incorrect |
| Model Development | Poorly | 2 | *7 | 2 | *7 | 2 | 7 | 2 | 7 | 2 | *7 |
| | Moderately | 10 | 6 | 7 | 9 | 10 | 6 | 10 | 6 | 10 | 6 |
| | Well | 25 | 10 | 29 | 6 | 28 | 7 | 28 | 7 | 25 | 10 |
| Internal Validation | Poorly | 1 | *8 | 2 | **7 | 1 | *8 | 2 | **7 | 1 | *8 |
| | Moderately | 10 | +6 | 7 | 9 | 9 | +7 | 5 | +11 | 8 | +8 |
| | Well | 25 | 10 | 23 | 12 | 27 | 8 | 25 | *10 | 24 | 11 |
| External Validation | Poorly | 2 | 0 | 1 | 1 | 2 | 0 | 1 | 1 | 2 | 0 |
| | Moderately | 1 | 1 | 0 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |
| | Well | 2 | 0 | 2 | 0 | 1 | 1 | 2 | 0 | 2 | 0 |

Table 10. Comparison of all five models made with different descriptor types and solvation methods for protein binding. Bold box indicates the model with the best predictions overall. Compounds in each absorption category in the training set (model development & internal validation): 27 low binding and 30 high binding. Compounds in each absorption category in the test set (external validation): 1 low binding and 3 high binding.

| | | Gas Phase | | | | | | Born Solvated | | | |
|---------------------|------|-----------|-----------|---------|-----------|---------|-----------|---------------|-----------|---------|-----------|
| | | 2D | | i3D | | 2D/3D | | i3D | | 2D/3D | |
| | | Correct | Incorrect | Correct | Incorrect | Correct | Incorrect | Correct | Incorrect | Correct | Incorrect |
| Model Development | Low | 27 | 0 | 25 | 2 | 26 | 1 | 25 | 2 | 27 | 0 |
| | High | 27 | 3 | 23 | 7 | 28 | 2 | 27 | 3 | 29 | 1 |
| Internal Validation | Low | 25 | 2 | 25 | 2 | 25 | 2 | 24 | 3 | 27 | 0 |
| | High | 26 | 4 | 20 | 10 | 26 | 4 | 22 | 8 | 24 | 6 |
| External Validation | Low | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| | High | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 |

When examining all the comparisons in Tables 8-10 of HIA, Caco-2 permeability, and protein binding together, there are subtle differences in accuracy of the models. Looking at the results in terms of quantitative data is the only way to determine how well the models performed. The models built using the 2D descriptors are very similar to the models built using the i3D descriptors computed for conformations modeled using Born solvation (with on average only 1 or 2 more compounds predicted in the incorrect category at each stage for the 2D descriptor model), although the i3D models all give slightly better results and mimic physiological conditions better. Using the i3D descriptor model with Born solvation requires a significantly greater amount of computational time and data storage capacity due to the necessity of the conformational search. If the goal of a specific project was to explore a small database of potential compounds, the i3D Born model may prove to be the best to use. However, if the goal of a specific project was to explore a large database of compounds, the 2D model may be the best approach to save the computational time and effort of finding the lowest energy conformations. The 2D descriptor model does not require a conformational search since the descriptors are not dependent upon the 3-dimensional structure of the molecule.

Model performance could also be impacted by compound stereochemistry, as conformational searches required input of a specific stereoisomer, although experimental data used to compile the database was not always attributed to a specific stereoisomer. Stereochemistry can affect the biological activity of a drug. It is possible for a drug that is chiral to be stereoselective for certain receptors, with this stereoselectivity varying for different receptors. Isomers can have a variety of effects—one isomer can be more potent than the other (called the eutomer and distomer respectively), both can be biologically active but only one leads to adverse effects (like toxicity), or one enantiomer may nullify out the effect of the other by antagonizing it.¹⁰³ Diastereomers are configured in three-dimensional space differently, being different compounds, which means their interactions with receptors will be different as well, leading to different biological effects. If a chiral drug has two enantiomers, they may differ in their bioavailability (which is a factor related to absorption), rate of metabolism, excretion, their selectivity for transporters (which affects the drug distribution), as well as toxicity.¹⁰⁴ The results of the models generated herein may be impacted by the assigned stereochemistry for some compounds, when a racemic mixture or different isomer may have been tested, due to these various stereochemical effects related to drug-receptor interaction, which can affect the absorption or distribution of a drug.

Commonly Mispredicted Compounds

In this work there was only one commonly mispredicted compound within all five models developed for protein binding. This compound was hydralazine (structure shown in Figure 13). A single compound does not allow determination of structural types that may not be accurately predicted when the models are applied to additional compounds.

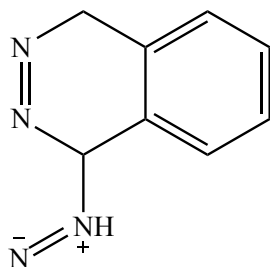


Figure 13. Structure of hydralazine. Only commonly mispredicted compound in the five protein binding models. Structure shown at physiological pH 7.4.

There were a total of seven commonly mispredicted compounds for the 5 HIA models, and the histogram of compound similarities can be seen in Figure 14. A table of all the mispredicted compounds within the HIA models is shown in the Appendix, Table A1.

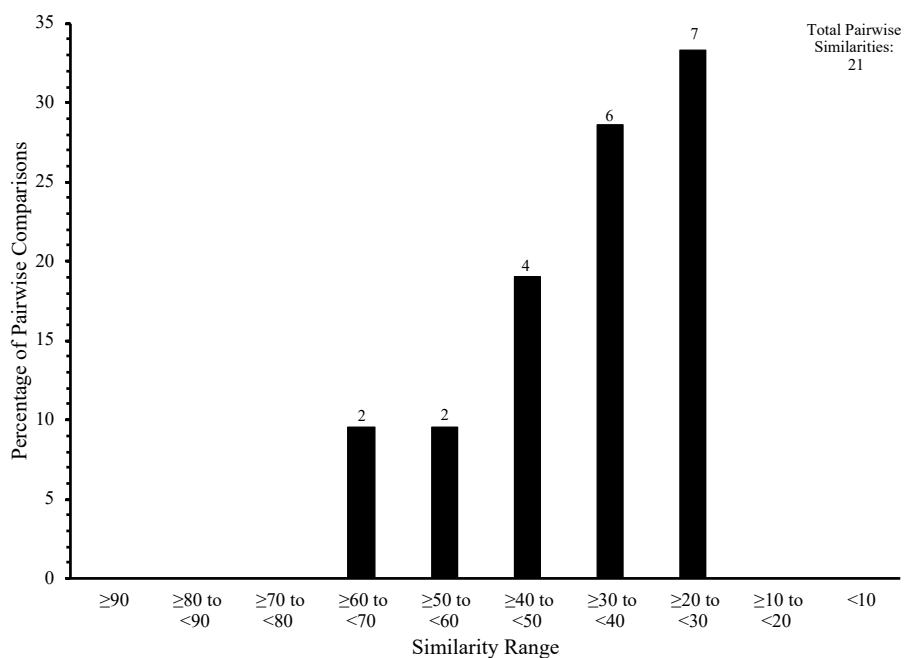


Figure 14. Histogram of the commonly mispredicted compounds within the HIA QSAR models. Seven total compounds were mispredicted. MACCS keys were calculated along with the Tanimoto coefficient to determine any compound chemical similarity. The majority of the pairwise similarities fall within the 20-40% bins.

The majority of the pairwise similarities in this group fell within the 20-40% range, suggesting limited structural similarity and providing no specific mechanism to identify, in advance, compounds that might be mispredicted when the model is applied to predict absorption properties of unknowns. The 5 Caco-2 permeability models had fifteen commonly mispredicted compounds and the resulting histogram can be seen in Figure 15. A table of all the mispredicted compounds within the Caco-2 permeability models is shown in the Appendix, Table A2.

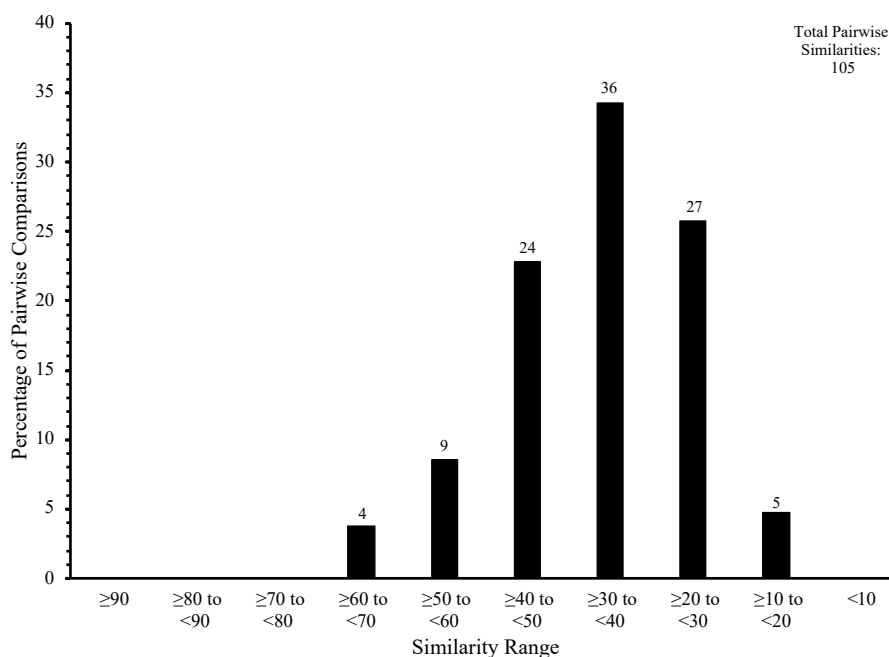


Figure 15. Histogram of the commonly mispredicted compounds within the Caco-2 permeability QSAR models. There was a total of fifteen commonly mispredicted compounds within the five descriptor model types. MACCS keys and Tanimoto coefficients were calculated to determine any compound chemical similarity. The majority of the pairwise similarities fall within the 20-50% range.

The majority of the pairwise similarities fell within the 20-50% range, once again showing they were structurally dissimilar. There was only one compound commonly

mispredicted between the five HIA models and the five Caco-2 models: tranexamic acid, which can be seen in Figure 16.

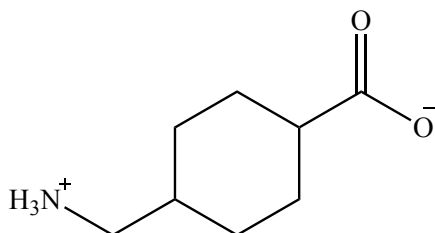


Figure 16. Structure of tranexamic acid. The only compound commonly mispredicted in all the HIA and Caco-2 permeability QSAR models. Structure shown at physiological pH 7.4.

Due to the low pairwise similarities for mispredictions made by the HIA and Caco-2 permeability models, the dataset was too structurally diverse to determine any potential ‘trouble’ compound structures to avoid in the future. This was indicated initially by the histograms created to determine the structural diversity of the dataset selected to be used in the models, but this exploration seemed imperative to determine whether any similarities occurred to help determine appropriate compounds on which to apply the models.

Conclusions

Overall, the models generated herein using different molecular descriptor types and solvation methods, all performed similarly for each descriptor type for HIA, Caco-2 permeability, and % protein binding. However, when examining the empirical breakdown for each, the i3D Born descriptor model types performed slightly better than the others. Descriptors in these models are computed for compounds under solvation conditions that mimic the environment of the human body. The HIA i3D Born model predicted 49 of 60 compounds correct in the model development stage, 43 of 60 correct in the internal validation stage & 4 of 6

correct in the external validation stage, with both well absorbed compounds correct and 1 moderate and poorly absorbed each correct. The Caco-2 permeability i3D Born model predicted 40 of 60 correct at the model development stage, 32 of 60 correct in the internal validation, and 4 of 6 correct in the external validation stage, with both well absorbed correct and 1 moderate and poorly absorbed each correct. The Caco-2 permeability models performed slightly worse than models for the other two activity types, at each stage of model development and validation, due to the large standard deviations on experimental data for some compounds in the reported datasets. The protein binding i3D Born model predicted 52 of 57 compounds correct at the model development stage, 46 of 57 correct at the internal validation, and 3 of 4 correct in the external validation stage, with the low binding compound correct and 2 of the high binding compounds correct.

Due to the similar behavior of all descriptor-type and solvation models, the application of these models in the future may be dependent upon the size of the datasets. Smaller datasets (less than 100 compounds) should use the i3D Born models, due to the slightly better predictions as discussed above and the mimicry of the body's natural environment. However, these descriptor-type models utilized a conformational search performed on the HPC, which requires computational time and is not ideal for larger datasets. Due to this, and the similar performance of the 2D descriptor type models, larger datasets (over 100 compounds) should use the 2D descriptor models. Unfortunately, we were unable to determine specific classes of compounds to be used to filter out 'troubling' compounds for future predictions, because our datasets were too structurally diverse. However, these models developed within this work can still be used to filter potential compounds for screening efforts.

Chapter 3

Method Development of Protein Binding Assays

Introduction

Protein binding (PB) measurements have previously been used to help determine the distribution of potential drugs in the human body. Proteins in blood plasma (albumin and α_1 -acid glycoprotein) bind drugs and effect distribution.¹⁰⁵ Some of the main methods previously utilized to investigate PB are equilibrium dialysis, ultracentrifugation, ultrafiltration, as well as gel filtration.¹⁰⁵ We chose not to use these specific approaches due to known disadvantages for each. Some disadvantages of equilibrium dialysis (the most widely used technique in this regard) are that it requires costly equipment (rapid equilibrium dialysis, RED), as well as it suffers from membrane binding and nonspecific binding.^{105, 106} The main disadvantages of ultracentrifugation are that there is a large amount of unbound drug in the sedimentation that results, along with the cost of the ultracentrifuge instrumentation. In addition, only a small number of samples can be analyzed at a time, and careful control of temperature and pH are needed.¹⁰⁵ Ultrafiltration is a relatively faster and easier technique than equilibrium dialysis, as well as providing accurate and quantitative data. However, there are issues with nonspecific binding and temperature and pH control are required.¹⁰⁵ Gel filtration does not have any membrane-related issues, however binding to the gel can occur, as well as being time consuming.¹⁰⁵

Herein we chose two spectroscopic techniques, namely, fluorescence polarization and nano differential scanning fluorimetry. Fluorescence polarization assays have also been used previously to determine protein binding.⁵⁵ Since its first application in 1952, this technique has been modified and its application expanded to be used in high-throughput screening (HTS) approaches. An advantage of FP in this analysis is the inclusion of internal quality control

parameters (ratio of emission intensities of bound and free probes) inherent in the technique.¹⁰⁷ Nano differential scanning fluorimetry is a technique typically used to determine protein stability via the spectrofluorometric determination of proteins.¹⁰⁸ However, it has also been used previously to determine yes/no protein binding checks.⁵⁶ These methods were selected for use herein, in part because of these strengths, their relative lack of weakness inherent in the other approaches, and also because instrumentation capable of performing these assays were readily available to us.

Fluorescence Polarization

Fluorescence polarization and fluorescence polarization anisotropy are assays used to determine binding between proteins (or other targets) and ligands, typically set up as competition assays where polarization of a fluorophore reporter should decrease as a small molecule competitor binds and displaces the reporter from its protein binding site.¹⁰⁹ Fluorescence polarization assays measure the change in polarized light of a fluorescent molecule as binding occurs by monitoring the emitted fluorescence light parallel and perpendicular to the excitation plane; as binding of the fluorescent reporter occurs, polarization increases due to a slower rotation of the reporter when complexed than when free in solution. Small molecules, like the fluorophore alone, rotate quickly in solution and so have low polarization as the ratio of the parallel and perpendicular fluorescence emitted is equal. Large molecules, like a protein, rotate more slowly in solution and thus induce high polarization upon binding due to the parallel fluorescence emitted being larger than the perpendicular fluorescence. Figure 17 shows an example of these changes in polarization caused by different molecule rotations.

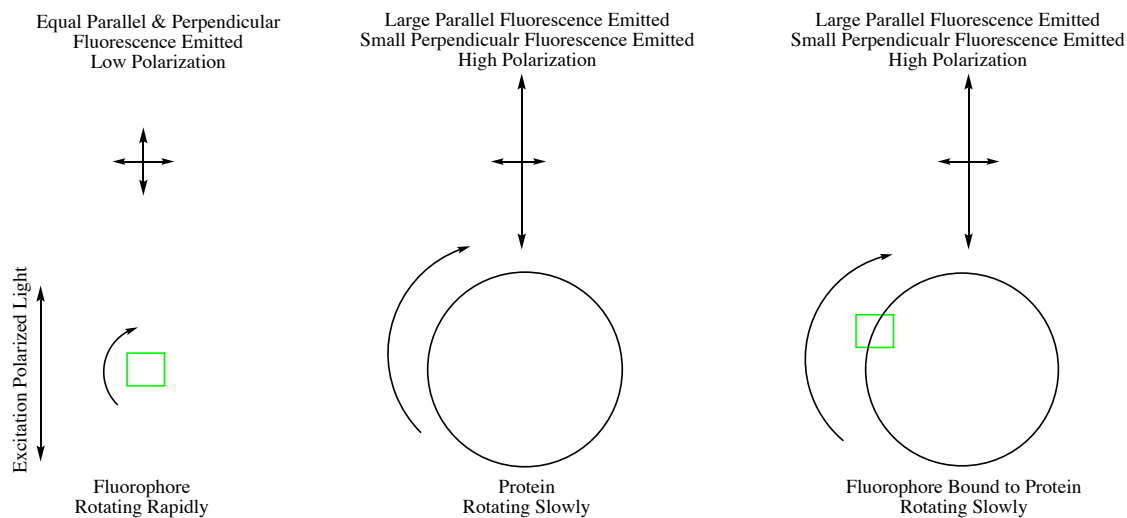


Figure 17. Schematic representation of fluorescence polarization. Smaller molecules (fluorophore) rotate rapidly in solution and produce low polarization values due to the parallel & perpendicular fluorescence emitted being equal (de-polarized). Larger molecules (protein) rotate slowly in solution and produce high polarization values due to the parallel fluorescence emitted being larger than the perpendicular fluorescence emitted (polarized). When a small molecule binds to a large molecule, the complex rotates slowly in solution and polarized light is emitted. Measuring this change from low polarization to high polarization determines protein binding.

The ratio of parallel and perpendicular fluorescent light emitted is used to determine the polarization (P) of molecules; most instruments output results as milliPolarization (mP) ($P \times 1000$) readouts along with the raw parallel (F_{\parallel}) and perpendicular (F_{\perp}) light emitted.¹¹⁰ Equation 4 shows the formula used to calculate milliPolarization by most instruments.

$$mP = 1000 * \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + F_{\perp}} \quad [\text{Eq. 4}]$$

The instrument used for these experiments was the FlexStation 3 from Molecular Devices. This is an automated multi-mode microplate reader capable of obtaining absorbance, fluorescence, luminescence, and fluorescence polarization measurements.¹¹¹ This is a fully

automated microplate reader, which is capable of quantitatively transferring solutions from one plate (reagent plate) into the assay plate with pipetter tips integral to the instrument, depending on how the protocol is set up in the SoftMax Pro software. The fluorescence polarization experiments described herein were implemented only after absorbance and emission experiments of the fluorophore in question were first analyzed to determine the parameters (excitation and emission wavelengths) to be used in subsequent polarization experiments. Two fluorophores were used in these fluorescence polarization experiments, fluorescein sodium salt and TopFluor® LysoPA (a BODIPY-labeled lysophosphatidic acid (LPA)). Fluorescein sodium salt is known to bind to BSA¹¹² as can be seen in Figure 18. Lysophosphatidic Acid (LPA) is also known to be associated with serum albumin.¹¹³ TopFluor LPA has been previously used to determine binding to the enzyme autotaxin (ATX) through the use of fluorescence anisotropy.¹¹⁴ The structure of TopFluor LPA, referred to as BODIPY-LPA throughout the rest of this dissertation, can be seen in Figure 19. These two fluorophores were selected for several reasons. The main reason being that they were both readily available. Fluorescein has been used for many years to determine binding to BSA,^{115, 116} it is relatively cheap,¹¹⁷ and has a relatively short fluorescence lifetime (~4 ns), which means fluorescence polarization is sufficiently high because fluorescence polarization decreases as excited state lifetimes increase.¹¹⁸ BODIPY-LPA has not been used in many previous fluorescence polarization experiments,^{114, 119} however LPA is known to bind to BSA, which is one of the main reasons we selected this fluorophore. BODIPY-LPA is also relatively expensive,¹²⁰ which is a downside to using this fluorescent-labeled lipid. BODIPY-LPA has a longer fluorescent lifetime than fluorescein (~6 ns), which means the fluorescence polarization is slightly lower than of fluorescein,¹¹⁸ which is another drawback to using this lipid.

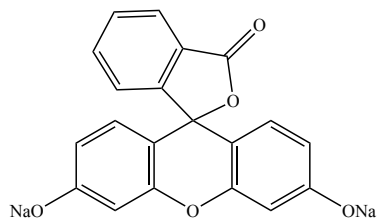


Figure 18. Structure of fluorescein sodium salt used in these experiments.

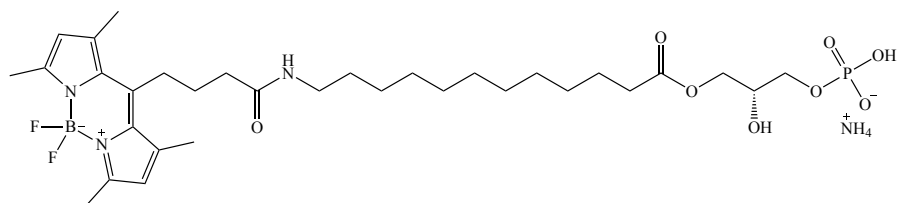


Figure 19. Structure of TopFluor LPA (BODIPY-LPA) used in these experiments.

Nano Differential Scanning Fluorimetry

Nano differential scanning fluorimetry (NanoDSF) was performed using the Tycho NT.6 (NanoTemper Technologies, Munich). This instrument can be used for rapid analysis of protein binding.⁵⁶ Nano differential scanning fluorimetry (NanoDSF) determines changes in the ratio of intrinsic fluorescence of tryptophan and tyrosine residues in a target protein to determine when unfolding of the protein occurs as temperature increases.^{56, 121} As the protein unfolds the intrinsic fluorescence ratio of tryptophan and tyrosine residues changes due to a change in environment associated with protein conformation as it unfolds. Figure 20 shows a general representation of how NanoDSF can be used to determine binding.

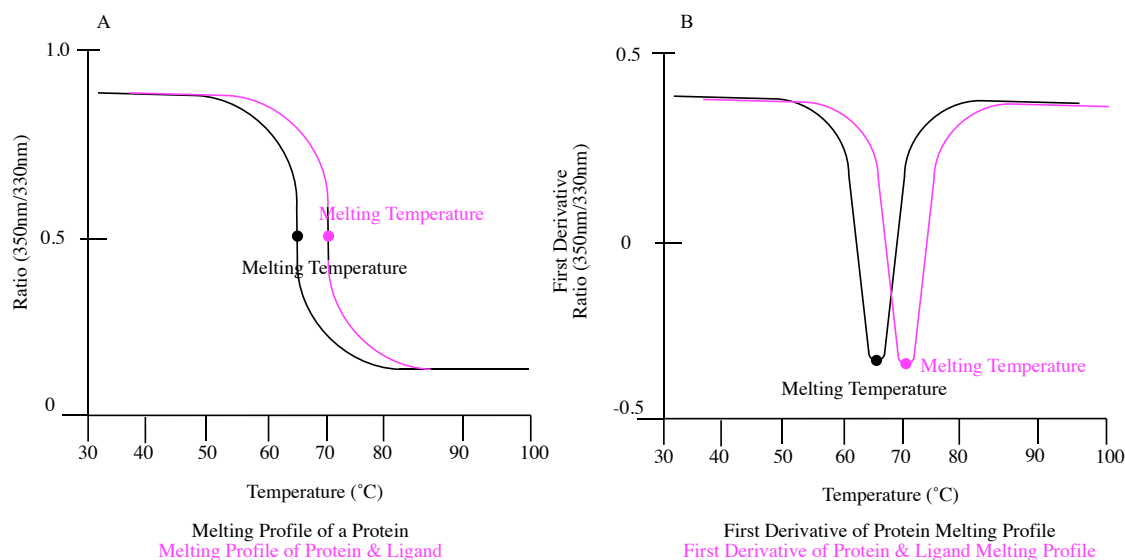
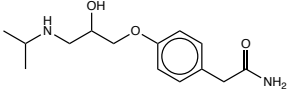
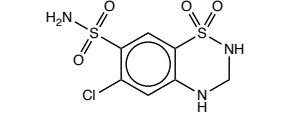
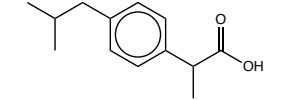
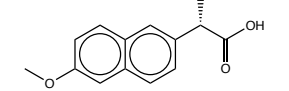


Figure 20. Schematic representation of Nano Differential Scanning Fluorimetry used to determine protein binding. As temperature increases, the ratio of intrinsic fluorescence of tryptophan (350nm) and tyrosine (330nm) residues changes, causing a melting profile, as seen in panel A. The first derivative of the melting profile is taken to determine the melting temperature of the protein, as seen in panel B. Black line is protein alone, pink is protein and ligand bound; melting temperature change is typically observed when binding occurs.

The idea was that as binding occurred between BSA and the drug, the melting temperature (T_m) or inflection temperature (T_i) would change depending on if the drug was a high (increase) or low binder (no change, or less change). Compounds were selected for validation experiments because they fit the absorption categories well (they were within the boundary cutoffs mentioned in Chapter 2 and they would all ionize well for analysis on the mass spectrometer in future work). Caco-2 permeability assays require quantitation of compounds that permeate the monolayer through mass spectrometry measurements or other suitable methods. All compounds selected to be used in protein binding experiments along with their protein binding value, HIA, and Caco-2 data can be seen in Table 11.

Table 11. Drugs selected to be examined using the FlexStation 3 and Tycho NT.6. Protein binding, HIA, Caco-2 permeability, & structures shown.

| Drug Name | % Protein Binding ¹²²⁻¹²⁶ | % HIA ⁸³ | Caco-2 Permeability (log scale) ^{85, 127} | Structure |
|---------------------|--------------------------------------|---------------------|--|---|
| Atenolol | 5-16% | 50% | -5.79 |  |
| Hydrochlorothiazide | 40-68% | 68% | -5.85 |  |
| Ibuprofen | 90-99% | 98% | -4.32 |  |
| Naproxen | 98-99% | 99% | -4.31 |  |

Methods

Fluorescence Polarization

Fluorescence polarization (FP) experiments were performed using the FlexStation 3 (Molecular Devices LLC, San Jose) to investigate protein binding of various drugs to BSA using two different fluorophores, fluorescein sodium salt (Fluka Analytical, Sigma Aldrich, St. Louis), and TopFluor® LysoPA (BODIPY-LPA, Avanti Polar Lipids, Alabaster). Proper conditions for fluorescence polarization experiments were determined from absorbance and emission

experiments of each fluorophore in buffer (1X PBS). Specifically, the excitation and emission wavelengths, as well as the cutoff wavelength were determined from these experiments.

Fluorescein Fluorescence Polarization Experiments

Absorbance and Emission Measurements of Fluorescein

A 10.0 mM stock solution of fluorescein was initially prepared in distilled water. This stock solution was then diluted for preliminary absorbance experiments. An intermediate stock of 50.0 μ M fluorescein was first prepared by adding 5 μ L of the 10.0 mM stock to 995 μ L 1X PBS (non-sterile, prepared in house). This 50.0 μ M intermediate stock was then used to prepare a 10.0 μ M working solution of fluorescein through the addition of 40 μ L (of the 50.0 μ M stock) into 160 μ L 1X PBS. This solution was prepared in the assay plate (Corning Costar tissue culture treated black, glass-bottom 96-well plate). An absorbance protocol for the FlexStation 3 was developed using the onboard SoftMax Pro software, where wavelengths read, how often the instrument read those wavelengths, what columns in the assay plate were read, and if any compound transfers occurred from the reagent plate were controlled. Table 12 shows the protocol used to obtain the absorbance spectrum of fluorescein.

Table 12. FlexStation 3 protocol created in the SoftMax Pro software used for absorbance measurements of fluorescein.

| Absorbance Spectrum Scan Read | |
|-------------------------------|------------|
| Wavelengths Read | 400-525 nm |
| Read Every | 10 nm |
| Reading Column | 3 |
| Compound Transfer | None |
| Temperature | 37 °C |

Fluorescence emission determination experiments for fluorescein were performed after determining the excitation wavelength of to be used (490nm) from the absorbance measurements. The 50 μM intermediate stock of fluorescein was used to prepare a 1.5 μM solution of fluorescein, diluted to a final concentration of 0.3 μM in the emission experiments. The 10.0 μM solution of fluorescein used in the absorbance experiments was too concentrated and gave saturated data. This 1.5 μM solution of fluorescein was prepared by adding 6 μL of the 50.0 μM intermediate stock to 194 μL of 1X PBS (non-sterile, prepared in house). The 1.5 μM solution was diluted to a 0.3 μM concentration when the instrument transferred 25 μL from the reagent plate (Corning Costar v-bottom, 300 μL 96-well microplate) into the 100 μL of buffer (1X PBS, pH 7.4, non-sterile, prepared in house) already in the read plate, so the dilution occurred in the instrument, eliminating human error. Table 13 shows the protocol used to obtain the fluorescence emission spectrum of fluorescein.

Table 13. FlexStation 3 protocol created in the SoftMax Pro software used for emission measurements of fluorescein.

| Fluorescence Emission Scan Read | |
|---------------------------------|-----------------|
| Excitation Wavelength | 490 nm |
| Emission Wavelengths Read | 450-550 nm |
| Read Every | 10 nm |
| Cutoff | 515 nm |
| PMT Gain | Medium |
| Flashes per Read | 6 |
| Reading Column | 6 |
| Compound Transfer | 1 |
| Reagent Column | 6 |
| Volume Transferred | 25 μ L |
| Volume in Read Plate | 100 μ L |
| Tip Used | 6 |
| Temperature | 37 $^{\circ}$ C |

Fluorescence Polarization Measurements of Fluorescein

Fluorescence polarization measurements of varying concentrations of fluorescein were performed to determine an optimal concentration range to use in later competition assays. A 100.0 μ M intermediate stock of fluorescein was prepared from the 10.0 mM stock in distilled water by adding 10 μ L of 10.0 mM to 990 μ L 1X PBS (non-sterile). This 100.0 μ M intermediate was then used to prepare the highest dose response concentration of 5.0 μ M by adding 75 μ L of

100.0 μM to 1425 μL 1X PBS (non-sterile). Serial dilutions were performed from this 5.0 μM solution to prepare the other dose response concentrations (2.5, 1.0, 0.5, 0.1, 0.05, and 0.01 μM). These solutions (200 μL each well) were then pipetted into columns 9-11 of the assay plate (Corning Costar tissue culture-treated black, glass-bottom 96-well plate). The fluorescence polarization protocol for the FlexStation 3 was constructed using the onboard SoftMax Pro software, where the excitation and emission wavelengths, along with the cutoff wavelength, PMT gain (how much light is emitted through the sample, to ensure saturation of signal does not occur), how many flashes of light per read occur, what columns on the assay plate are read, and if any solution is transferred from the reagent plate are all controlled. Table 14 shows the protocol used to obtain these measurements.

Table 14. FlexStation 3 protocol created in the SoftMax Pro software used for fluorescence polarization measurements of fluorescein.

| Fluorescence Polarization Endpoint Read | |
|---|--------|
| Excitation Wavelength | 490 nm |
| Emission Wavelength | 525 nm |
| Cutoff | 515 nm |
| PMT Gain | Low |
| Flashes per Read | 100 |
| Reading Columns | 9-11 |
| Compound Transfer | None |
| Temperature | 37 °C |

Fluorescein Binding to Varying Concentrations of BSA

Binding of varying concentrations BSA and 2.5 μM fluorescein was examined in terms of fluorescence polarization. A 100.0 μM intermediate stock of fluorescein was prepared from the 10.0 mM stock in distilled water by adding 35 μL of 10.0 mM to 3465 μL Gibco Dulbecco's 1X PBS (DPBS, ThermoFisher Scientific, Waltham). This intermediate stock was then used to prepare the working solution of 50.0 μM fluorescein by adding 50.0 μL of the 100.0 μM to 50.0 μL DPBS. This working solution was prepared in the reagent plate (Corning Costar V-bottom, 300 μL 96-well microplate). This solution was 20X concentrated to dilute to 2.5 μM in the read (assay) plate when 10 μL was added by the instrument. A 111.0 μM BSA solution was prepared by dissolving 40.96 mg BSA into 5555 μL DPBS and two serial dilutions occurred to prepare the other two concentrations of BSA (11.1 and 1.11 μM). These solutions were 1.11X concentrated to dilute to 100.0, 10.0, and 1.0 μM BSA when the fluorescein was added by the instrument. These solutions were pipetted into the read plate (Corning Costar non-tissue culture-treated, black flat-bottom 96-well), so there were 3 replicates of each concentration, and the FlexStation protocol was constructed in the onboard SoftMax Pro software. The solutions were read every 59 seconds for 30 minutes to ensure that equilibrium binding had occurred. These experiments were done to determine the best concentration of BSA to use with 2.5 μM fluorescein in the competition assays.

Competition Assays of BSA Binding (Fluorescein and Drugs)

Competition binding assays between fluorescein and varying drug concentrations were based on methods outlined by Mathias & Jung in 2007.¹²⁸ These studies used human serum albumin (HSA) at 10.0 μM with a fluorophore and a naproxen dose response ranging from 0.01

μM to 10 mM. Our competition assays were designed similarly with 10.0 μM BSA, 2.5 μM fluorescein, and a drug dose response of 0, 1, 10, 100, 1000, & 10000 μM . The drugs used in these competition assays were atenolol, hydrochlorothiazide, naproxen, and ibuprofen (a low binder, mid-binder, and two high binders, respectively). A 50.0 μM working solution of fluorescein was prepared from the 10.0 mM stock in distilled water by adding 10 μL into 1990 μL DPBS. This working solution was pipetted into a column of the reagent plate, a Corning Costar V-bottom 300 μL 96-well plate (150 μL each well). This solution was 20X concentrated so a concentration of 2.5 μM fluorescein was in the final assay. A 13.3 μM BSA solution was prepared by dissolving 3.53 mg BSA into 4 mL DPBS. This solution was pipetted into 3 columns of the read plate, rows A-G, so 150 μL were in each well. Row H had 150 μL DPBS pipetted into 3 columns, so a fluorescein blank could be measured to normalize data. The read plate was a Corning Costar black, non-tissue culture-treated flat-bottom 96-well microplate. Drug solutions 5X concentrated were prepared in DMSO by dissolving each drug in 1 mL of DMSO to prepare a 50,000 μM solution. From this solution, serial dilutions were used to prepare the other dose response concentrations. These solutions were pipetted into a column on the reagent plate (200 μL rows A-F, 200 μL DMSO in rows G & H). The FlexStation protocol was set up in the SoftMax Pro software so 10 μL of the fluorescein was pipetted by the instrument into 180 μL BSA first and read for 30 minutes. Then, 40 μL of the drug solutions were pipetted by the instrument into the 190 μL BSA/fluorescein and read for a total of 4 hours. This means each well had a total of 20% final DMSO concentration. Triplicate experiments of each drug were performed, means & standard deviations calculated, and the fluorescence polarization data was normalized and scaled from 0-100 as shown in Equation 5, where x is the original mP value, *min*

is the mP of 2.5 μ M fluorescein in DPBS, and *max* is the mP of 10.0 μ M BSA & 2.5 μ M fluorescein.

$$\% mP = \frac{(x-min)}{(max-min)} * 100 \quad [\text{Eq. 5}]$$

Calculations of % protein binding are based on an equation outlined by Zhang, Nordeen, and Shapiro in 2013, who applied a fluorescence anisotropy method to determine binding between a steroid receptor activator and steroid receptors.¹²⁹ The equation used in our methods is modified from this paper and is shown in Equation 6, where the % protein binding is determined at a specific concentration, the mP change is calculated from the difference between mP when no fluorophore is present and the mP in the presence of the competitor (drug), and the max. mP change is calculated from the largest change in mP independent of which concentration is being examined, following the same guidelines as the mP change in the numerator.

$$\% \textit{protein binding} = \frac{\textit{mP change}}{\textit{max. mP change}} * 100 \quad [\text{Eq.6}]$$

However, one thing to note is that Equation 5 scales all mP graphs in these competition assays to 100%, so this Equation 5 is inherently calculating these changes in mP used for Equation 6. Our graphs also show the % protein binding of the fluorophore, so the opposite value is used for the competitor (drug). For example, if 90% mP is shown on the graph, then 10% mP is used for the competitor, and therefore, 10% of the drug is bound.

BODIPY-LPA Fluorescence Polarization Experiments

Absorbance and Emission Measurements of BODIPY-LPA

A stock solution of 1 mg/mL BODIPY-LPA was prepared by dissolving 1 mg BODIPY-LPA in 1 mL chloroform. This stock solution was aliquoted into 70 μ L volumes and dried with air in a fume hood. This 70 μ L aliquot was then resuspended in 1 mL DMSO to prepare a 100.0 μ M intermediate stock of BODIPY-LPA. From this 100.0 μ M intermediate stock, a 5.0 μ M solution was prepared by adding 75 μ L of 100.0 μ M to 1425 μ L 1X PBS (non-sterile). This 5.0 μ M solution (200 μ L each well) was pipetted into the assay plate (Corning Costar tissue culture-treated black, glass-bottom 96-well plate). The absorbance protocol for the FlexStation 3 was compiled using the onboard SoftMax Pro software, where wavelengths read, how often those wavelengths were read, what columns of the assay plate were read, and if any solution was transferred from the reagent plate was controlled. Table 15 shows the protocol used to obtain the absorbance spectrum of BODIPY-LPA.

Table 15. FlexStation 3 protocol created in the SoftMax Pro software used for absorbance measurements of BODIPY-LPA.

| Absorbance Spectrum Scan Read | |
|-------------------------------|------------|
| Wavelengths Read | 400-600 nm |
| Read Every | 10 nm |
| Reading Column | 1 |
| Compound Transfer | None |
| Temperature | 37 °C |

Emission experiments for BODIPY-LPA were performed after determining the excitation wavelength (500nm) to be used from the absorbance measurements. The same 5.0 μ M BODIPY-LPA solution prepared for the absorbance measurements was pipetted into different columns on the same assay plate and the fluorescence emission was measured using the SoftMax Pro protocol shown in Table 16, where excitation wavelength, what wavelengths were being read and how often they were read, what cutoff wavelength was set, as well as PMT gain, and how many flashes of light occurred per read, what column of the assay plate was read, and if any solution was transferred from the reagent plate were all controlled.

Table 16. FlexStation 3 protocol created in the SoftMax Pro software used for emission measurements of BODIPY-LPA.

| Fluorescence Emission Scan Read | |
|---------------------------------|------------|
| Excitation Wavelength | 500 nm |
| Emission Wavelengths Read | 400-700 nm |
| Read Every | 10 nm |
| Cutoff | 530 nm |
| PMT Gain | Medium |
| Flashes per Read | 6 |
| Reading Column | 2 |
| Compound Transfer | None |
| Temperature | 37 °C |

Fluorescence Polarization Measurements of BODIPY-LPA

Fluorescence polarization measurements of varying concentrations of BODIPY-LPA were performed to determine the optimal concentration to use in later competition assays. A 100.0 μM intermediate stock was prepared by resuspending a frozen, dried, 70 μL aliquot (1 mg/ml) in 1 mL DMSO. This intermediate stock was used to prepare a 5.0 μM solution by adding 75 μL of 100.0 μM to 1425 μL 1X PBS (non-sterile). Serial dilutions were then performed to prepare the other dose response concentrations (2.5, 1.0, 0.5, 0.1, 0.05, and 0.01 μM). These solutions (200 μL each well) were pipetted into columns 6-8 of the assay plate (Corning Costar tissue culture-treated, black glass-bottom 96-well microplate). The fluorescence polarization protocol for the FlexStation 3 was compiled using the onboard SoftMax Pro software, where excitation, emission, and cutoff wavelengths were set, PMT gain and how many flashes of light occurred per read were set, and what columns were read on the assay plate, along with if any solution transferred from the reagent plate were all controlled. Table 17 shows the protocol used to obtain these measurements.

Table 17. FlexStation 3 protocol created in the SoftMax Pro software used for fluorescence polarization measurements of BODIPY-LPA.

| Fluorescence Polarization Endpoint Read | |
|---|--------|
| Excitation Wavelength | 500 nm |
| Emission Wavelength | 540 nm |
| Cutoff | 530 nm |
| PMT Gain | Medium |
| Flashes per Read | 100 |
| Reading Columns | 6-8 |
| Compound Transfer | None |
| Temperature | 37 °C |

BODIPY-LPA Binding to Varying Concentrations of BSA

Binding of varying concentrations BSA and 1.0 μM BODIPY-LPA was examined in terms of fluorescence polarization. A 100.0 μM intermediate stock of BODIPY-LPA was prepared by resuspending a 70 μL aliquot of 1 mg/mL (dried down, frozen) in 1 mL DMSO. From this intermediate stock, the working solution of 20.0 μM was prepared by adding 20.0 μL 100.0 μM to 80.0 μL DPBS. This solution was prepared in the reagent plate (Corning Costar V-bottom 300 μL 96-well plate). This solution was 20X concentrated to dilute to 2.5 μM in the read (assay) plate when 10 μL aliquots were added by the instrument. A 111.0 μM BSA solution was prepared by dissolving 40.96 mg BSA into 5555 μL DPBS and two serial dilutions occurred to prepare the other two concentrations of BSA (11.1 and 1.11 μM). These solutions were 1.11X concentrated to dilute down to 100.0, 10.0, and 1.0 μM BSA when the BODIPY-LPA was added

by the instrument. These solutions were pipetted into the read plate (Corning Costar non-tissue culture-treated, black flat-bottom 96-well), so there were 3 replicates of each concentration, and the FlexStation protocol was set up in the SoftMax Pro software. The solutions were read for 30 minutes to ensure binding occurred. These experiments were done to determine the best concentration of BSA to use with 1.0 μM BODIPY-LPA in the competition assays.

Competition Assays of BSA Binding (BODIPY-LPA and Drugs)

Competition binding assays between BODIPY-LPA and varying competing drug concentrations were based on methods outlined by Mathias & Jung in 2007.¹²⁸ Our competition assays were designed similarly with 10.0 μM BSA, 1.0 μM BODIPY-LPA, and a drug dose response of 0, 1, 10, 100, 1000, & 10000 μM . The drugs used in these competition assays were atenolol, hydrochlorothiazide, naproxen, and ibuprofen (the same low, mid-, and high binders as used in the fluorescein experiments). A 20.0 μM working solution of BODIPY-LPA was prepared from the 100.0 μM intermediate stock prepared from a 70 μL aliquot of 1 mg/mL by adding 400 μL of 100.0 μM to 1600 μL DPBS. This working solution was pipetted into a column of the reagent plate, a Corning Costar V-bottom 300 μL 96-well plate (150 μL each well). This solution was 20X concentrated so a concentration of 1.0 μM BODIPY-LPA was in the final assay. A 13.3 μM BSA solution was prepared by dissolving 3.53 mg BSA into 4 mL DPBS. This solution was pipetted into 3 columns of the read plate, rows A-G, so 150 μL were in each well. Row H had 150 μL DPBS pipetted into 3 columns, so a BODIPY-LPA blank could be measured to normalize data. The read plate was a Corning Costar black, non-tissue culture-treated flat-bottom 96-well microplate. Drug solutions 5X concentrated were prepared in DMSO by dissolving each drug in 1 mL of DMSO to prepare a 50,000 μM solution. From this solution,

serial dilutions were used to prepare the other dose response concentrations. These solutions were pipetted into a column on the reagent plate (200 μ L rows A-F, 200 μ L DMSO in rows G & H). The FlexStation protocol was compiled using the onboard SoftMax Pro software so 10 μ L of the BODIPY-LPA was pipetted by the instrument into 180 μ L BSA first and read for 30 minutes. Then, 40 μ L of the drug solutions were pipetted by the instrument into the 190 μ L BSA/BODIPY-LPA and read for a total of 4 hours. This means each well had a total of 21% final DMSO concentration. Triplicate experiments of each drug were performed, means & standard deviations calculated, and the fluorescence polarization data was normalized and scaled from 0-100 as shown in Equation 5, where x is the original mP value, min is the mP of 1.0 μ M BODIPY-LPA in DPBS, and max is the mP of 10.0 μ M BSA & 1.0 μ M BODIPY-LPA. Calculations of % protein binding are based on Equation 6 and described in the competition assay of fluorescein section of this dissertation.

Nano Differential Scanning Fluorimetry

Nano differential scanning fluorimetry (NanoDSF) was performed using the Tycho NT.6 (NanoTemper Technologies, Munich) to investigate an alternative, fast, and inexpensive assay to assess protein binding of various drugs to bovine serum albumin (BSA, Sigma Aldrich, St. Louis). In each experiment, the fluorescence ratio of the emission wavelengths 350 nm and 330 nm was monitored as a function of temperature in the range of 35 $^{\circ}$ C to 95 $^{\circ}$ C. Protein fluorescence is typically measured at absorbance maximums near 280 nm (used as an excitation wavelength).¹³⁰ This wavelength is used because the aromatic residues in proteins (tryptophan, tyrosine, and phenylalanine) have absorbance maximums at or near 280nm.¹³¹

Control Assays

Control experiments were conducted using 100 μM concentrations of each drug target in buffer to ensure no interference in the fluorescent signal. Samples for control experiments were prepared by adding 10 μL of the 5.0 mM drug stock in DMSO to 490 μL of buffer (1 X PBS, nonsterile) to make a 100 μM solution, with 2% DMSO. Samples were performed in 1.5 microfuge tubes. Duplicate experiments (3 replicates each) were performed. All drugs were purchased from Sigma Aldrich, St. Louis and were used without purification.

Binding Assays

Binding experiments were performed to determine if a change in melting temperature of 10.0 μM BSA occurred with a dose response (0, 100, 1000, and 10000 μM) of each drug (atenolol, hydrochlorothiazide, and ibuprofen). Samples for binding experiments were prepared by adding 200 μL of each intermediate stock of drug in DMSO (500, 5000, and 50000 μM) or DMSO vehicle to 800 μL 12.5 μM BSA in Gibco Dulbecco's 1X PBS (DPBS, ThermoFisher Scientific, Waltham), with 20% DMSO. The 12.5 μM BSA solution was prepared by dissolving 9.96 mg BSA (lyophilized powder, $\geq 96\%$, Sigma Aldrich, St. Louis) in 12 mL DPBS. This BSA solution was then split into 800 μL in each of the twelve 1.5 mL microfuge tubes (3 replicates triplicate experiments). After 200 μL each intermediate stock of drug was added to the BSA solution, the samples were incubated at 37 $^{\circ}\text{C}$ for 4 hours to ensure equilibrium binding was reached. After incubation, triplicate experiments were performed on the Tycho NT.6. The change in melting temperature was then determined using the first derivative of the protein profile (change in ratio as temperature increased).

Results and Discussion

Fluorescence Polarization

Fluorescein Experiments

Absorbance and Emission Spectra for Fluorescein

The absorbance and emission spectra of fluorescein in buffer (1X PBS, pH 7.4) was determined using the FlexStation 3. A 10 μM solution of fluorescein under these conditions had an absorption maximum at 490 nm. This maximum was then used to excite a 0.3 μM fluorescein solution to determine the emission spectrum. A fluorescence maximum of 525 nm was observed under these conditions, and a cutoff of 515 nm was found to yield the best results (eliminating the absorbance spectra, so only the emission maximum was observed). The absorbance and emission scans for fluorescein under these conditions can be seen in Figure 21. These excitation and emission wavelengths were used in all further experiments with fluorescein as fluorophore.

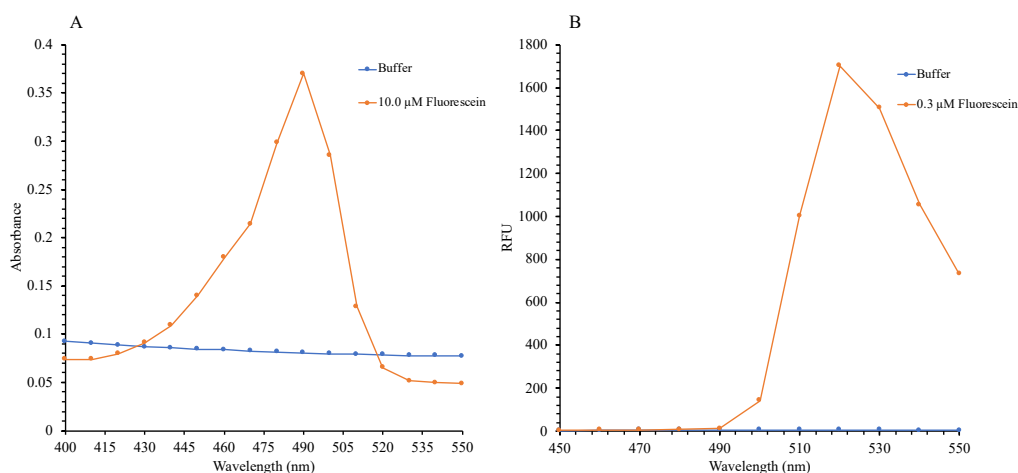


Figure 21. Absorbance and Emission spectra of fluorescein sodium salt and buffer (1X PBS, pH 7.4). Panel A is the absorbance of 10.0 μM fluorescein in buffer (1X PBS, pH 7.4). The absorbance maximum under these conditions was determined to be 490 nm. Panel B is the emission of 0.3 μM fluorescein in the same buffer using 490 nm as excitation. For all subsequent experimental work with fluorescein, these conditions were utilized (excitation 490 nm, emission 525 nm with a cutoff of 515 nm).

Fluorescence Polarization of Fluorescein

A dose response of fluorescein in buffer (1X PBS, pH 7.4) was examined to determine the best concentration to use for future BSA competition binding experiments. In buffered solutions containing only fluorescein, emissions show an even distribution between parallel fluorescence and perpendicular fluorescence, indicating the molecule is freely and rapidly rotating in solution rapidly (Figure 20). As the concentration of fluorescein increases, the fluorescence polarization (milliPolarization, mP) decreases. This is due to the fluorescence emitted increasing, which causes the ratio used to calculate mP to decrease. These data can be seen in Figure 22. These data also indicate that fluorescein at 2.5 μM produces a relatively high signal (RFU and measurable mP) near the top of the linear portion of the dose response curve, while maintaining requiring a modest amount of reporter. This concentration was used in all further experiments.

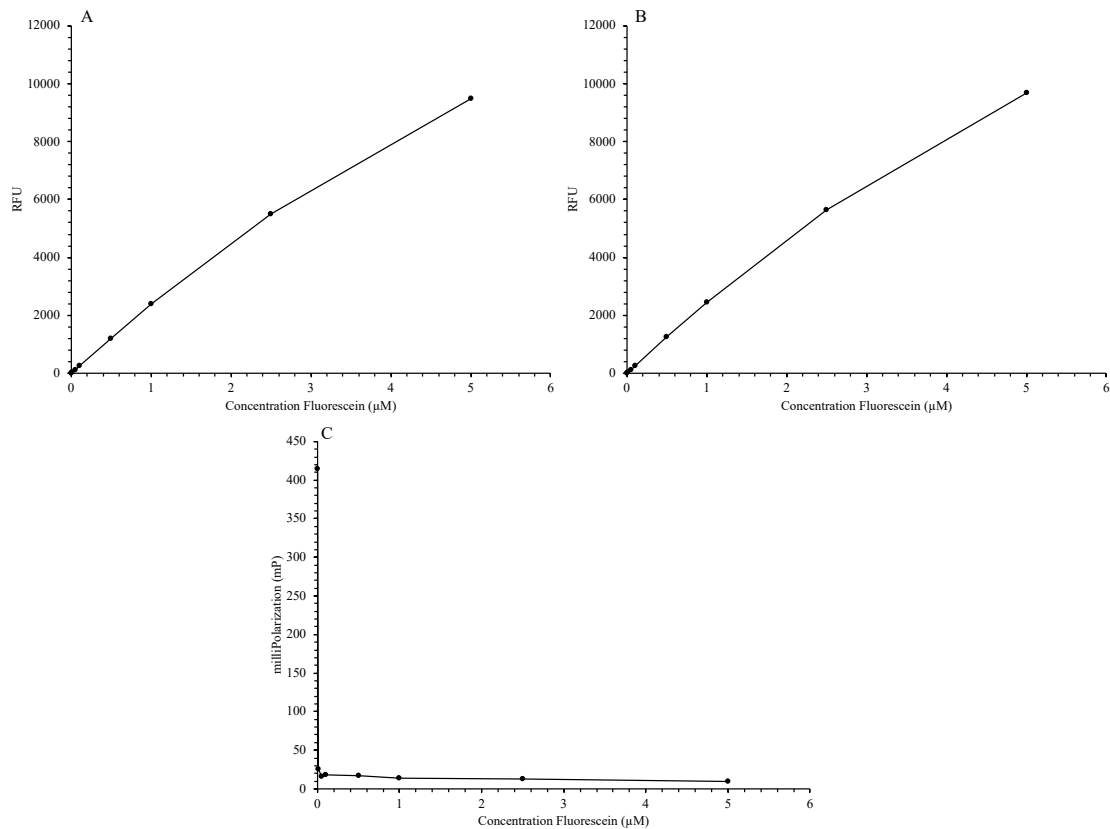


Figure 22. Perpendicular (panel A) and parallel (panel B) emission spectrum of dose response of fluorescein in buffer (1X PBS, pH 7.4). Fluorescence polarization of dose response of fluorescein in buffer (1X PBS, pH 7.4) (panel C). Concentrations of fluorescein are 0.01, 0.05, 0.1, 0.5, 1.0, 2.5, and 5.0 µM. Excitation 490nm, Emission 525nm, cutoff 515nm.

Fluorescein Binding to Varying Concentrations of BSA

Binding experiments between 2.5 µM fluorescein and varying concentrations of BSA were performed to determine a baseline of binding between the fluorophore and BSA in terms of mP. The results of this experiment can be seen in Figure 23. The measured mP increases as concentration of BSA increases (with a fixed concentration of fluorescein), due to the molecules (BSA and fluorescein bound together) slowly rotating in solution causing the polarized light to

be emitted more from the parallel plane than the perpendicular plane. The binding was examined at varying timepoints to examine the time course of the effect (Figure 23). There was not a significant difference in mP between the 15-minute and 30-minute timepoints.

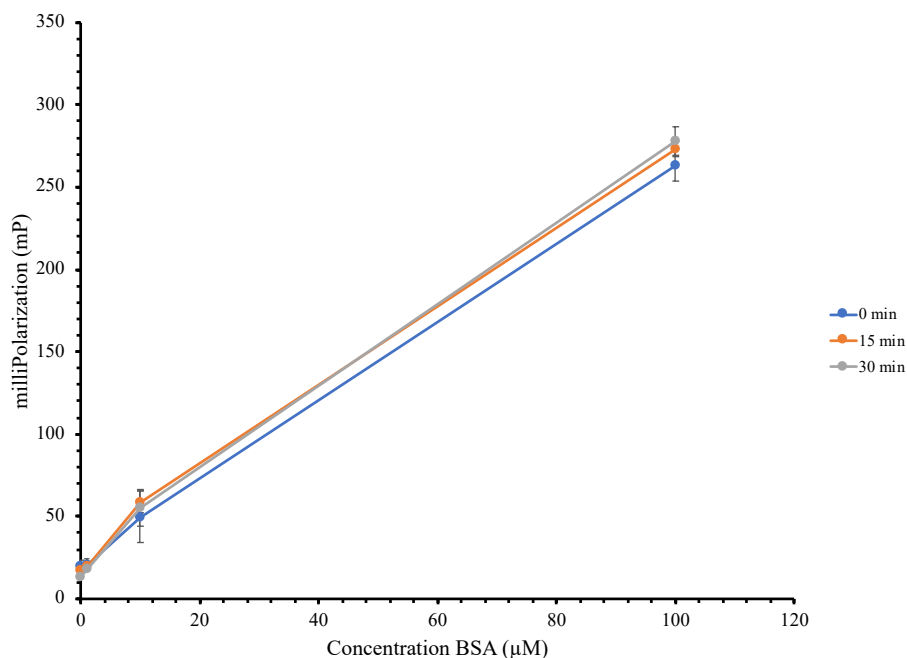


Figure 23. Fluorescence polarization (milliPolarization) of varying BSA concentrations (0, 1.0, 10.0, & 100.0 μM) in the presence of 2.5 μM fluorescein. Varying timepoints shown to determine binding occurred between fluorescein & BSA.

Competition Binding Assays of BSA, Fluorescein, and Drugs

Competition assays examining the binding of fluorescein (2.5 μM) and multiple drugs to BSA (10.0 μM) were performed. Figure 24 shows the competition between fluorescein (2.5 μM) and a dose response of atenolol, a low binder to blood proteins reported to be 5-16%.¹²² The observed decrease in mP is slight, which indicates that the fluorescein reporter was not competed off its BSA binding site by added drug. In short, atenolol did not displace the fluorescein reporter

from BSA. The decrease in mP indicates that only ~20% of atenolol is bound at the highest concentration (10 mM); this is based on the 80% mP shown on the graph, meaning 80% of the fluorescein is bound to BSA, and 20% of atenolol is bound (based on equations 5 & 6). This corresponds with atenolol being a low binder to BSA.

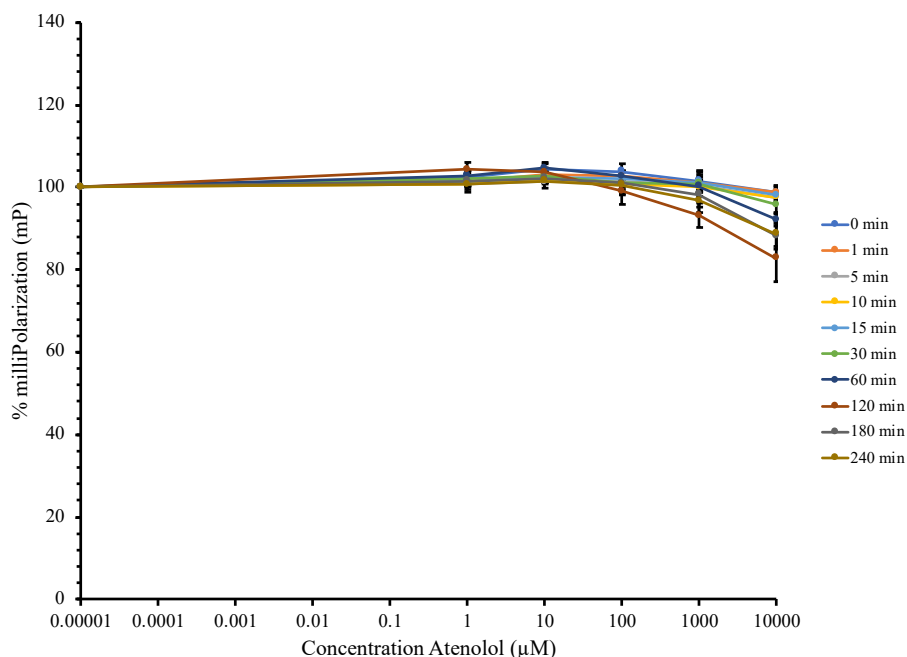


Figure 24. Average milliPolarization of atenolol dose response to 10.0 µM BSA and 2.5 µM fluorescein. Atenolol is a low binder with % protein binding 5-16%. Averages of triplicate experiments. Standard deviations shown as error bars. Values were normalized by subtracting fluorescein alone and scaled from 0-100 by dividing by BSA and fluorescein alone minus fluorescein alone and multiplying by 100 to get percent.

Figure 25 shows the competition between fluorescein (2.5 µM) and a dose response of hydrochlorothiazide, a mid-binder to blood proteins, reported to be 40-68%.¹²³ The decrease in mP is similar to that determined for atenolol, Again, hydrochlorothiazide did not displace fluorescein from its BSA binding site. The decrease in mP indicates that only ~20% of hydrochlorothiazide is bound at the highest concentration (10 mM); this is based on 80%

fluorescein bound (80% mP), which means 20% of hydrochlorothiazide is bound (based on equations 5 & 6). This does not match the 40-68% found in the literature, but it does indicate that low or mid-binders can still be distinguished by a slight decrease in mP, rather than a large decrease in mP.

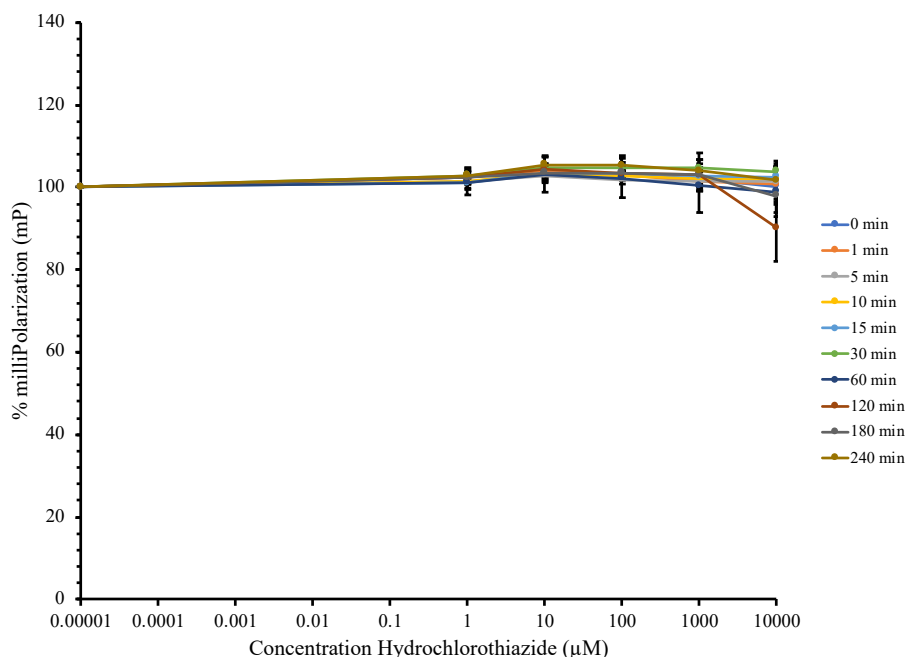


Figure 25. Average milliPolarization of hydrochlorothiazide dose response to 10.0 μM BSA and 2.5 μM fluorescein. Hydrochlorothiazide is a mid-binder with % protein binding 40-68% bound. Averages of triplicate experiments. Standard deviations shown as error bars. Values were normalized by subtracting fluorescein alone and scaled from 0-100 by dividing by BSA and fluorescein alone minus fluorescein alone and multiplying by 100 to get percent.

Figure 26 shows the competition between fluorescein (2.5 μM) and a dose response of naproxen, a high binder to blood proteins reported to be 98-99%.^{125, 126} The decrease in mP seen here is larger than that of the low & mid-binders, which indicates that the fluorescein was competed from its BSA binding site to a much higher degree than the previous examples. This

decrease in mP indicates that 80% of naproxen is bound at the highest concentration (10 mM); this is based on the 20% mP shown on the graph, which means 80% naproxen is bound (based on equations 5 & 6). This is less than the literature reports, but it does indicate that a difference can be seen between low and high binders using this method.

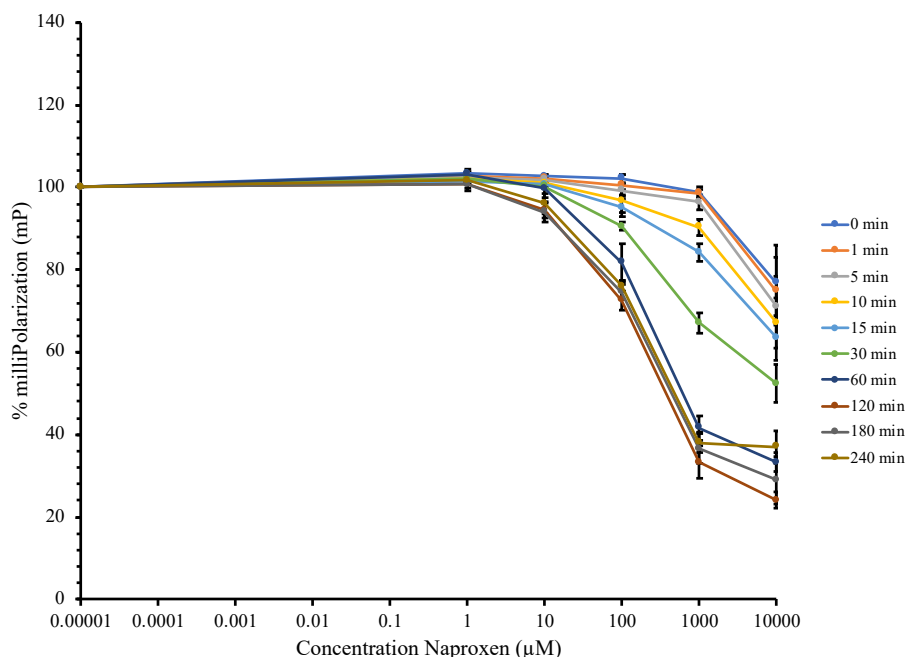


Figure 26. Average milliPolarization of naproxen dose response to 10.0 μM BSA and 2.5 μM fluorescein. Naproxen is a high binder with % protein binding 98-99% bound. Averages of triplicate experiments. Standard deviations shown as error bars. Values were normalized by subtracting fluorescein alone and scaled from 0-100 by dividing by BSA and fluorescein alone minus fluorescein alone and multiplying by 100 to get percent.

Figure 27 shows the competition between fluorescein (2.5 μM) and a dose response of ibuprofen, a high binder to blood proteins reported to be 90-99%.¹²⁴ The decrease in mP seen here is larger than that of the low & mid-binders, but still less than the naproxen (high binder) dose response. This indicates that the fluorescein is competed its BSA binding site, but not as much as the naproxen. This decrease in mP indicates that 45% of ibuprofen is bound at the

highest concentration (10 mM); this is based on the 55% mP shown in the graph, meaning 45% ibuprofen is bound (based on equations 5 & 6). However, at the 2-, 3-, and 4-hour timepoints, the resulting mP increases due to a significant decrease in the fluorescence. This is thought to be caused by aggregation over time which would decrease the fluorescence when the aggregates form enough to block the light. This was confirmed in nanoDSF experiments shown later, where a cloudy solution was noted after a 4-hour incubation of the highest ibuprofen concentration in a BSA solution.

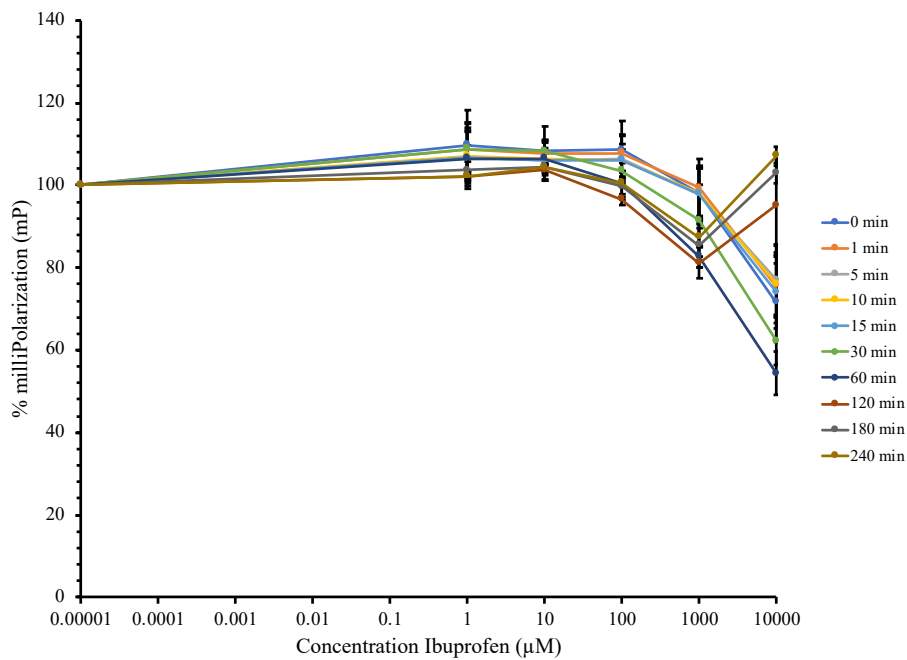


Figure 27. Average milliPolarization of ibuprofen dose response to 10.0 µM BSA and 2.5 µM fluorescein. Ibuprofen is a high binder with % protein binding 90-99% bound. Averages of triplicate experiments. Standard deviations shown as error bars. Values were normalized by subtracting fluorescein alone and scaled from 0-100 by dividing by BSA and fluorescein alone minus fluorescein alone and multiplying by 100 to get percent. At the highest concentration ibuprofen, an increase in mP is seen at 2 hours due to aggregation.

BODIPY-LPA Experiments

Absorbance and Emission Spectra for BODIPY-LPA

The absorbance and emission spectra of BODIPY-LPA in buffer (1X PBS, pH 7.4) was determined using the FlexStation 3. A 5.0 μM solution of BODIPY-LPA was found to have an absorbance maximum at 500 nm. This maximum was then used to excite a 5.0 μM BODIPY-LPA solution to determine the emission spectrum. A fluorescence maximum was seen at 540 nm under these conditions and a cutoff of 530 nm was found to yield the best results showing this maximum. Figure 28 shows the absorbance and emission spectrums of 5.0 μM BODIPY-LPA. These conditions were used in all further fluorescence polarization experiments of BODIPY-LPA.

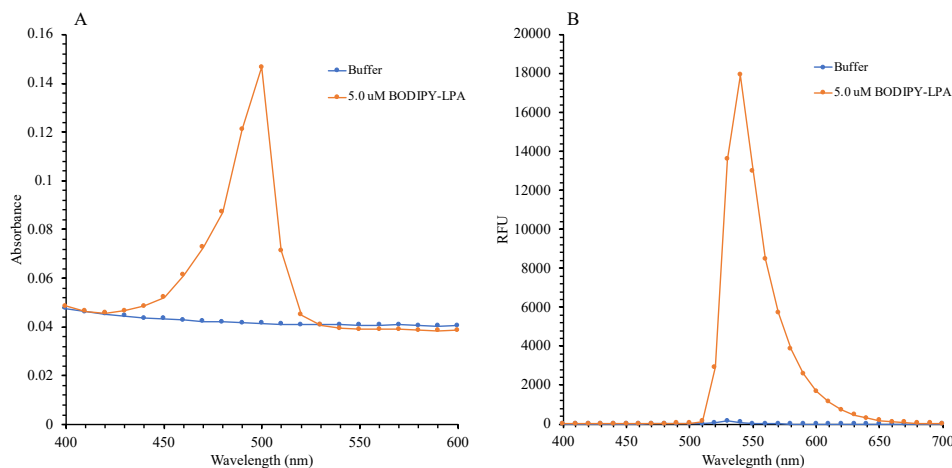


Figure 28. Absorbance and Emission spectra of BODIPY-LPA and buffer (1X PBS, pH 7.4). Panel A is the absorbance of 5.0 μM BODIPY-LPA in buffer (1X PBS, pH 7.4). The absorbance maximum under these conditions was determined to be 500 nm. Panel B is the emission of 5.0 μM BODIPY-LPA in the same buffer using 500 nm as excitation. For all subsequent experimental work with BODIPY-LPA, these conditions were utilized (excitation 500 nm, emission at 540 nm with a cutoff of 530 nm).

Fluorescence Polarization of BODIPY-LPA

A dose response of BODIPY-LPA in buffer (1X PBS, pH 7.4) was examined to determine the best concentration to use for future experiments in BSA completion assays. The parallel and perpendicular fluorescence showed an even distribution, indicating the molecule was rotating rapidly in solutions containing only reporter fluorophore (figure 29). The fluorescence increased as concentration of BODIPY-LPA increased, which is why the calculated mP decreased as concentration increased (the ratio of used to calculate mP is dependent upon fluorescence values from parallel and perpendicular light). This data is shown in Figure 29. The data shown here indicates that the concentration of BODIPY-LPA that produces a relatively high signal (RFU and measurable mP) and consistent mP is that for 1.0 μ M BODIPY-LPA. This concentration was used in all further experiments.

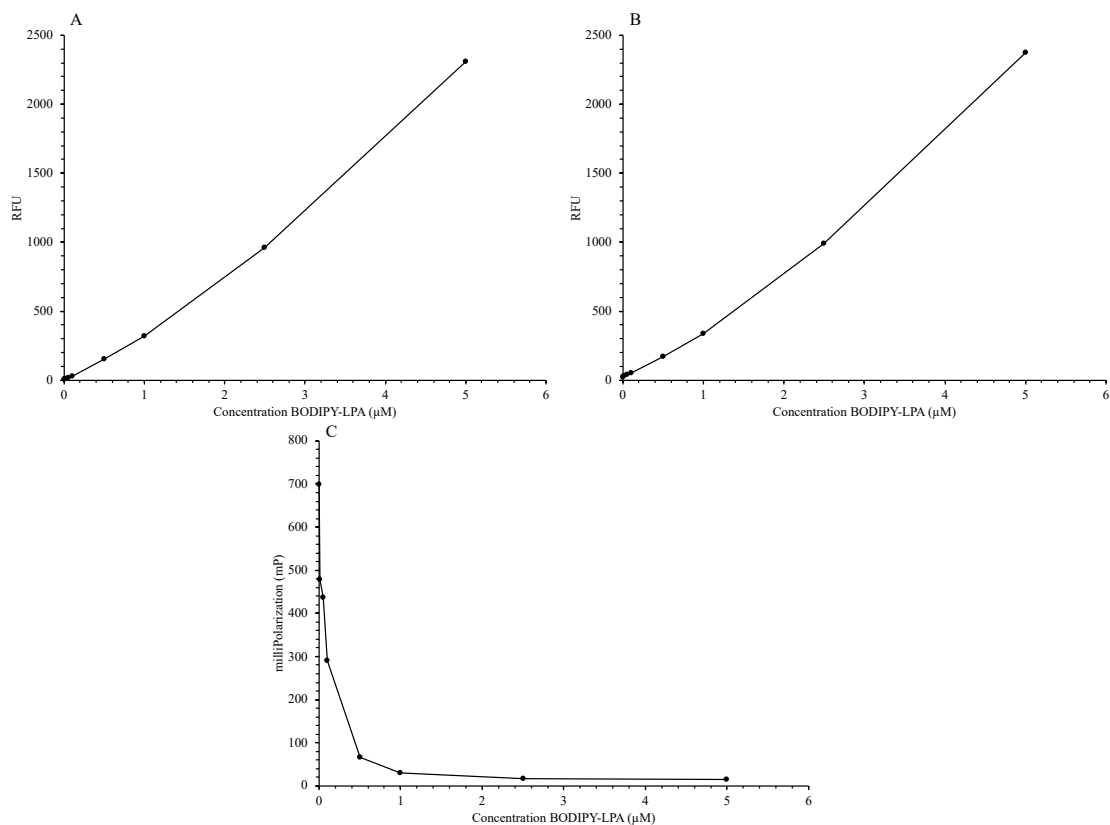


Figure 29. Perpendicular (panel A) and parallel (panel B) emission spectrum of dose response of BODIPY-LPA in buffer (1X PBS). Fluorescence polarization of dose response of BODIPY-LPA in buffer (1XPBS) (panel C). Concentrations of BODIPY-LPA are 0.01, 0.05, 0.1, 0.5, 1.0, 2.5, and 5.0 μM. Excitation 500nm, Emission 540nm, cutoff 530nm.

BODIPY-LPA Binding to Varying Concentrations of BSA

Binding experiments between 1.0 μM BODIPY-LPA and varying concentrations of BSA were performed to determine a baseline of binding between the reporter fluorophore and BSA measured as mP. This data can be seen in Figure 30. In these experiments mP increases as the concentration of BSA increases, due to the complex of BSA and fluorescein slowly rotating in solution which causes the polarized light to be emitted more from the parallel plane than the

perpendicular plane. Binding of the reporter fluorophore to BSA was examined at varying timepoints to examine the timecourse of change in mP. Figure 30 shows only a slight difference in mP between the 15-minute and 30-minute timepoints, indicating that binding likely maximizes within 15 minutes. This is similar to the binding seen between BSA & fluorescein, but the difference between the 0- and 15-minute timepoints is larger than was observed with the fluorescein (see Figure 23). The calculated mP resulting from binding between BODIPY-LPA and BSA is also higher than that for fluorescein and BSA. This is due to the raw fluorescence for BODIPY-LPA being lower than that of fluorescein.

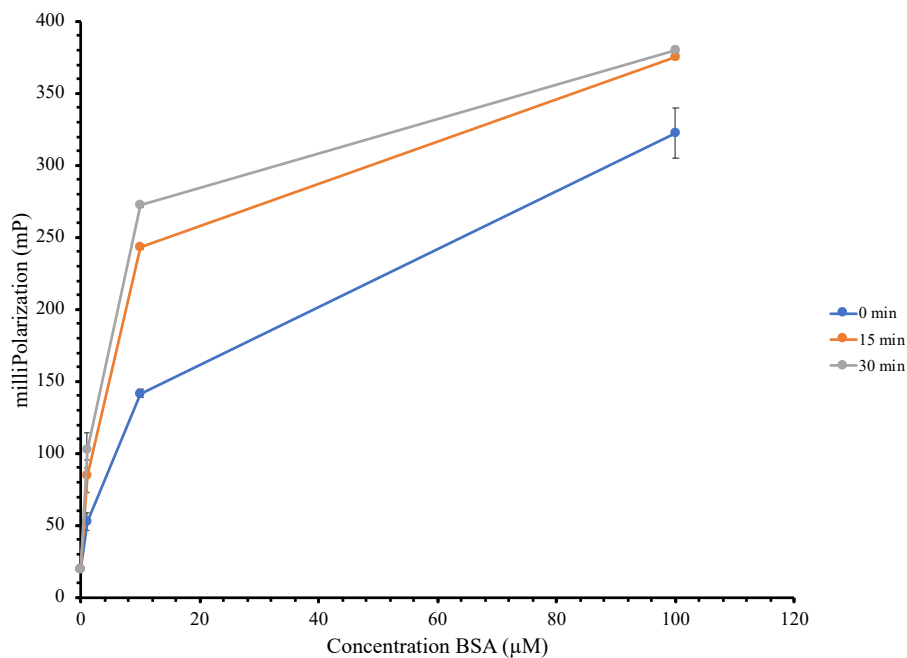


Figure 30. Fluorescence polarization of varying BSA concentrations (0, 1.0, 10.0, & 100.0 µM) in the presence of 1.0 µM BODIPY-LPA. Varying timepoints shown to determine binding occurred between BODIPY-LPA & BSA.

Competition Binding Assays of BSA, BODIPY-LPA, & Drugs

Competition assays between BODIPY-LPA (1.0 μM) and multiple drugs binding to BSA (10.0 μM) were performed. Figure 31 shows the competition between BODIPY-LPA (1.0 μM) and a dose response of atenolol, a low binder to blood proteins reported to be 5-16%. A slight increase in mP is seen as the concentration atenolol increases, opposite of what is observed with fluorescein. This could be because BODIPY-LPA is a tighter binder to BSA than fluorescein and is therefore more difficult to compete off. However, this still implies that this method can be used to distinguish between a low and high protein binder because a large decrease was not seen.

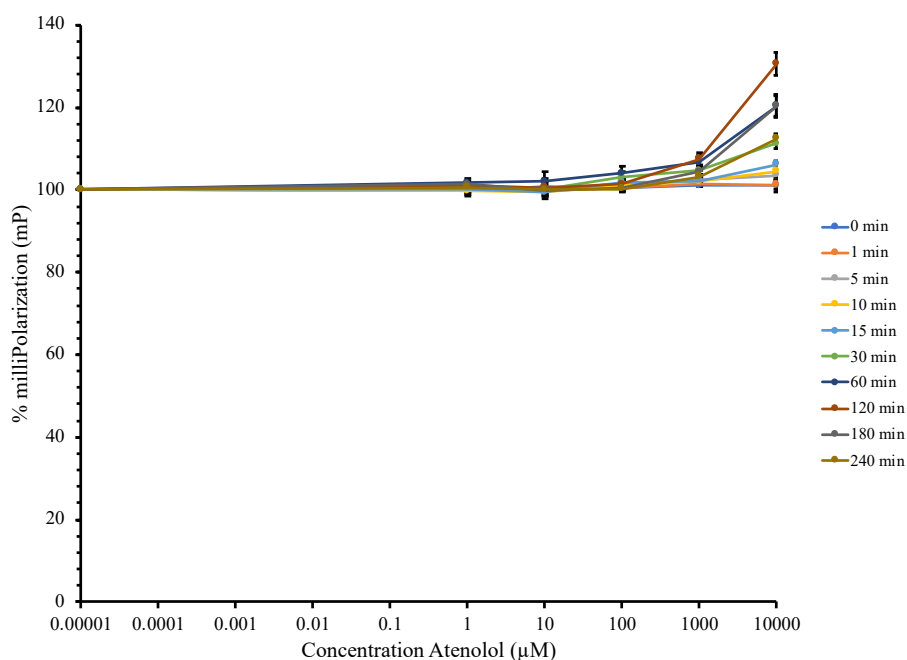


Figure 31. Average milliPolarization of atenolol dose response to 10.0 μM BSA and 1.0 μM BODIPY-LPA. Atenolol is a low binder with % protein binding 5-16%. Averages of triplicate experiments. Standard deviations shown as error bars. Values were normalized by subtracting BODIPY-LPA alone and scaled from 0-100 by dividing by BSA and BODIPY-LPA alone minus BODIPY-LPA alone and multiplying by 100 to get percent.

Figure 32 shows the competition between BODIPY-LPA (1.0 μM) and a dose response of hydrochlorothiazide, a mid-binder to blood proteins reported to be 40-68%. A slight decrease in mP is seen as the concentration of hydrochlorothiazide increases. This implies that $\sim 15\%$ of hydrochlorothiazide is bound at the highest concentration (10 mM); this is based on the 85% mP shown in the graph, indicating 15% hydrochlorothiazide is bound (based on equations 5 & 6), which is definitely lower than the literature reports. However, it still indicates that the method distinguishes between a low and high protein binder.

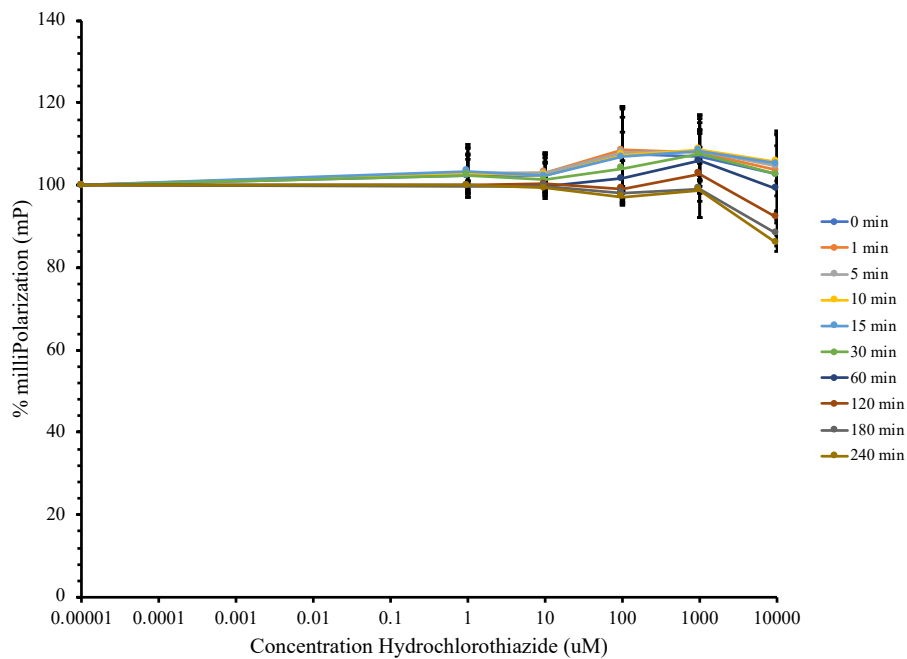


Figure 32. Average milliPolarization of hydrochlorothiazide dose response to 10.0 μM BSA and 1.0 μM BODIPY-LPA. Hydrochlorothiazide is a mid-binder with % protein binding 40-68% bound. Averages of triplicate experiments. Standard deviations shown as error bars. Values were normalized by subtracting BODIPY-LPA alone and scaled from 0-100 by dividing by BSA and BODIPY-LPA alone minus BODIPY-LPA alone and multiplying by 100 to get percent.

Figure 33 shows the competition between BODIPY-LPA (1.0 μM) and a dose response of naproxen, a high binder to blood proteins reported to be 98-99%. A large decrease in mP is seen as the concentration of naproxen increases. This implies that $\sim 65\%$ of naproxen is bound to BSA at the highest concentration (10 mM); this is based on the 35% mP shown on the graph, indicating 65% naproxen is bound (based on equations 5 & 6). This is a lower than reported in the literature, but it is still a significant change when compared to the low & mid-binders examined. The BODIPY-LPA is not competed from its BSA binding site by naproxen as much as fluorescein was, which further agrees with the idea that BODIPY-LPA is a tighter binder to BSA than fluorescein. However, these results further again show that this method can be used to distinguish between high and low protein binders.

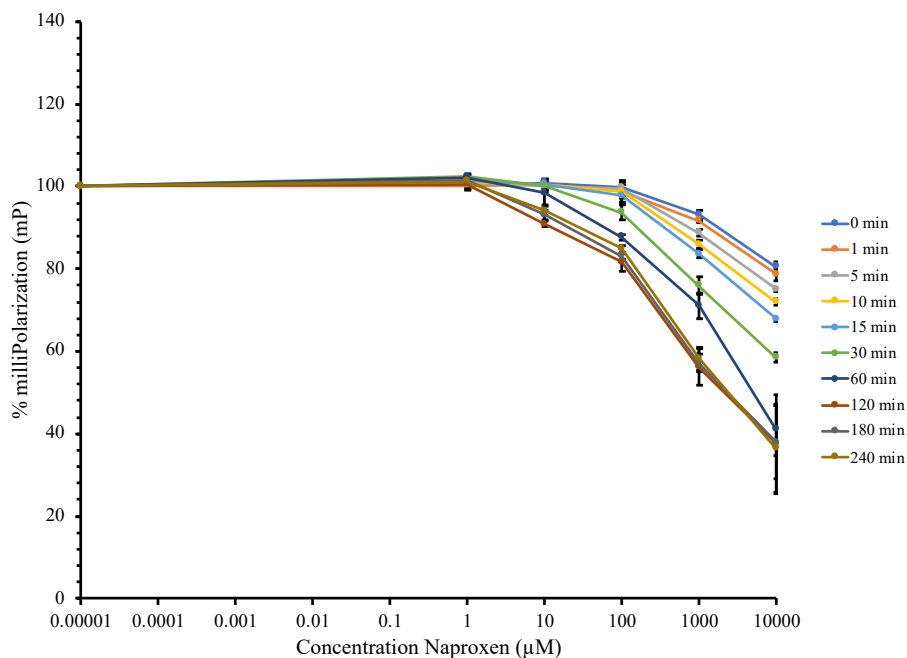


Figure 33. Average milliPolarization of naproxen dose response to 10.0 μM BSA and 1.0 μM BODIPY-LPA. Naproxen is a high binder with % protein binding 98-99% bound. Averages of triplicate experiments. Standard deviations shown as error bars. Values were normalized by subtracting BODIPY-LPA alone and scaled from 0-100 by dividing by BSA and BODIPY-LPA alone minus BODIPY-LPA alone and multiplying by 100 to get percent.

Figure 34 shows the competition between BODIPY-LPA (1.0 μM) and a dose response of ibuprofen, a high binder to blood proteins reported to be between 90-99%. At the highest concentration of ibuprofen (10 mM), an increase in mP is seen after 1 hour due to a decrease in fluorescence. This decrease in fluorescence is thought to be caused by aggregation, which is further supported by the nanoDSF data shown later in this dissertation when a cloudy solution was observed after incubation of 10 mM ibuprofen with BSA for 4 hours. Due to these significant increases in mP (up to 204% mP), the 2-, 3-, and 4-hour timepoint data were removed from the graph to keep scaling consistent. Due to this aggregation, the 1.0 mM concentration of ibuprofen data is what was used to determine that around 50% of ibuprofen was bound at the 4-

hour timepoint; this is based on 50% mP seen on the graph, indicating that the other 50% of BSA is occupied by ibuprofen (based on equations 5 & 6). Even though aggregation occurs, the change in mP is still a larger decrease than observed with the low and mid-binder, which indicates that this method can be used to determine a difference between low and high protein binders in the future.

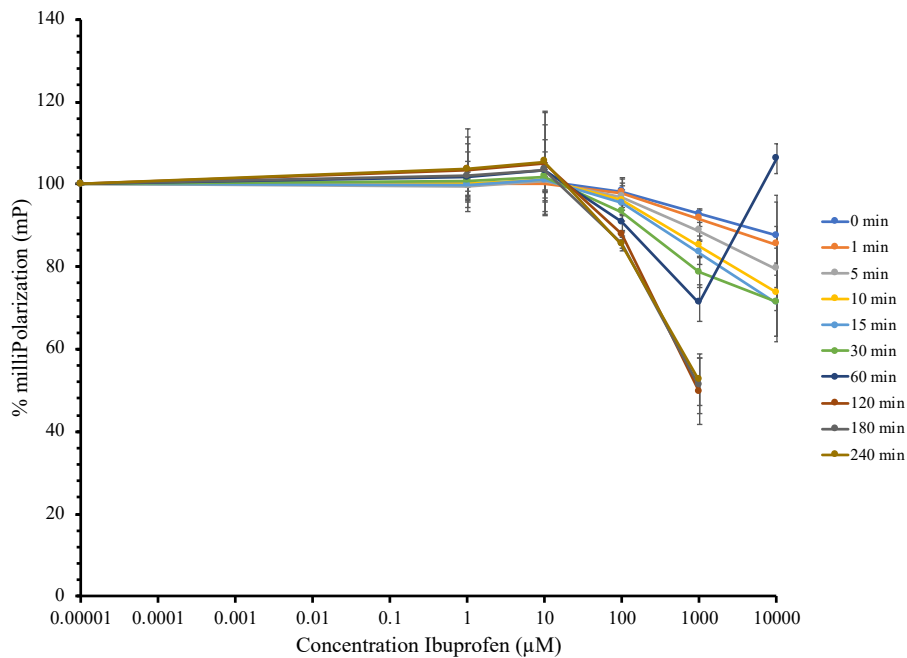


Figure 34. Average milliPolarization of ibuprofen dose response to 10.0 µM BSA and 1.0 µM BODIPY-LPA. Ibuprofen is a high binder with % protein binding 90-99% bound. Averages of triplicate experiments. Standard deviations shown as error bars. Values were normalized by subtracting BODIPY-LPA alone and scaled from 0-100 by dividing by BSA and BODIPY-LPA alone minus BODIPY-LPA alone and multiplying by 100 to get percent. At the highest concentration ibuprofen, an increase in mP is seen at 1 hour due to aggregation. The highest concentration ibuprofen data at 2, 3, & 4 hours removed due to a large increase in mP (up to 204 % mP); removed to keep scale consistent.

Nano Differential Scanning Fluorimetry

Control Assays

Control experiments for the NanoDSF methodology were performed to ensure that no interference occurred at the fluorescence signals measured (350nm/330nm). Figure 35 shows the results of these control experiments. The change in ratio (350nm/330nm) as a function of temperature was examined for 100.0 μ M of each drug (atenolol, hydrochlorothiazide, naproxen, and ibuprofen). Most of the compounds do not have interference at those wavelengths, with a ratio around 1.0 (atenolol, hydrochlorothiazide, and ibuprofen). Naproxen, however, was found to have a ratio around 7.0 as temperature increased, which implies it would cause interference in the binding measurements performed with BSA. As such, naproxen was not used in subsequent binding experiments.

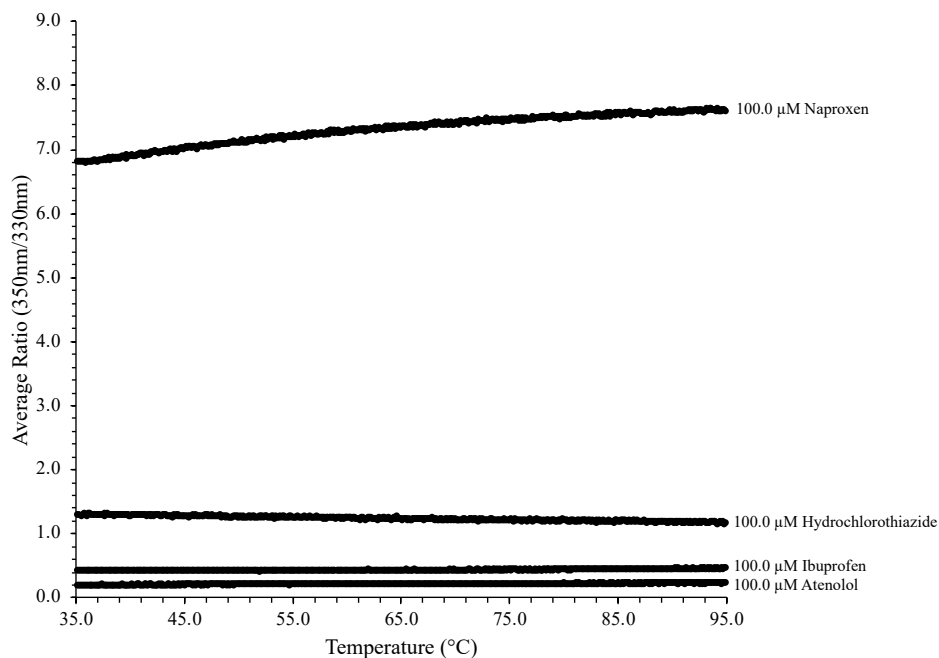


Figure 35. Average change in ratio (350nm/330nm) as a function of temperature for each compound. Most compounds have no change in fluorescence ratio as a function of temperature and low 350nm/330nm ratios around 1.0. Naproxen has an increase in fluorescence ratio as a function of temperature and a ratio around 7.0. Averages are of duplicate experiments.

Binding Assays

Binding experiments examined by NanoDSF were performed to determine whether the melting temperature (T_m) of BSA changes as different protein binders are introduced (low, mid, high). Figure 36 shows the melting temperature of 10.0 μM BSA in the presence of multiple concentrations of atenolol. The T_m of BSA did not change in the presence of different concentrations atenolol. A slight decrease in T_m is seen as concentration atenolol increases to 10 mM, from 58 °C (BSA alone) to 56 °C (10 mM atenolol). This decrease is small (only 2 degrees), as compared to tighter binders (see below).

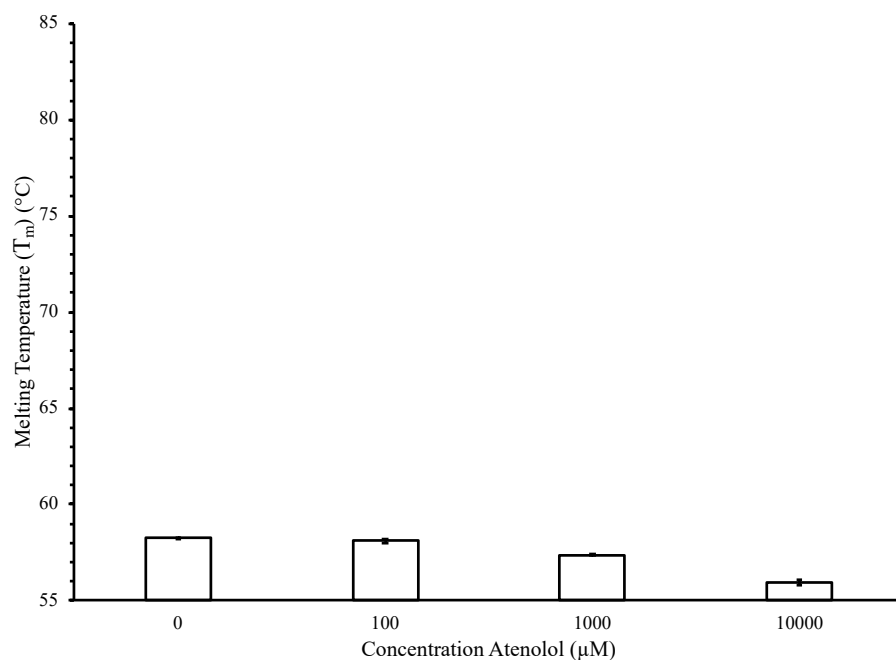


Figure 36. Average T_m of 10.0 μM BSA in the presence of varying concentrations of the drug atenolol. Atenolol is a low binder, with % protein binding around 5-16% bound. Averages of triplicate experiments. Standard deviations shown as error bars.

Figure 37 shows the melting temperature of 10.0 μM BSA in the presence of multiple concentrations of hydrochlorothiazide. The T_m of BSA does not change when in the presence of different concentrations of hydrochlorothiazide. The T_m of BSA alone is 58 °C and the T_m in the presence of 10 mM hydrochlorothiazide is nearly unchanged at 58 °C.

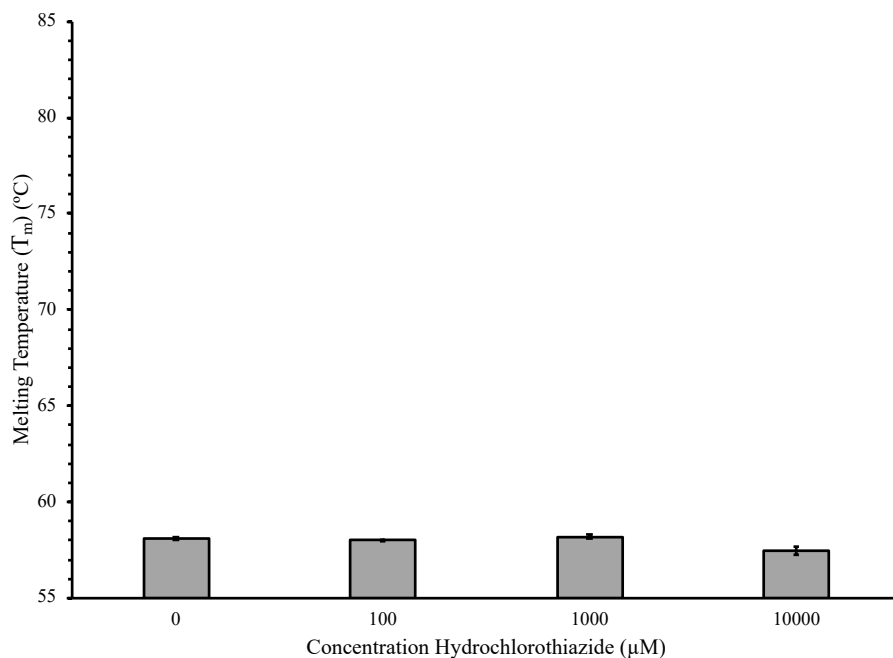


Figure 37. Average T_m of 10.0 μM BSA in the presence of varying concentrations of the drug hydrochlorothiazide. Hydrochlorothiazide is a mid-binder, with % protein binding around 40-68% bound. Averages of triplicate experiments. Standard deviations shown as error bars.

Figure 38 shows the melting temperature of 10.0 μM BSA in the presence of multiple concentrations of ibuprofen. The T_m of BSA increases as concentration of ibuprofen increases. The T_m of BSA increases from 58 °C (BSA alone) to 73 °C (10 mM ibuprofen). However, when performing these experiments, it was noted that the solution containing the 10 mM ibuprofen was cloudy, likely due to aggregation occurring in solution over time. With this being the case, when examining the change in T_m of BSA from 58 °C (BSA alone) to 70 °C (1 mM ibuprofen), the T_m still increases as concentration of ibuprofen increases. This data suggests that this method may be used to determine the difference between high and low binders by examining an increase in T_m in the presence of high binders.

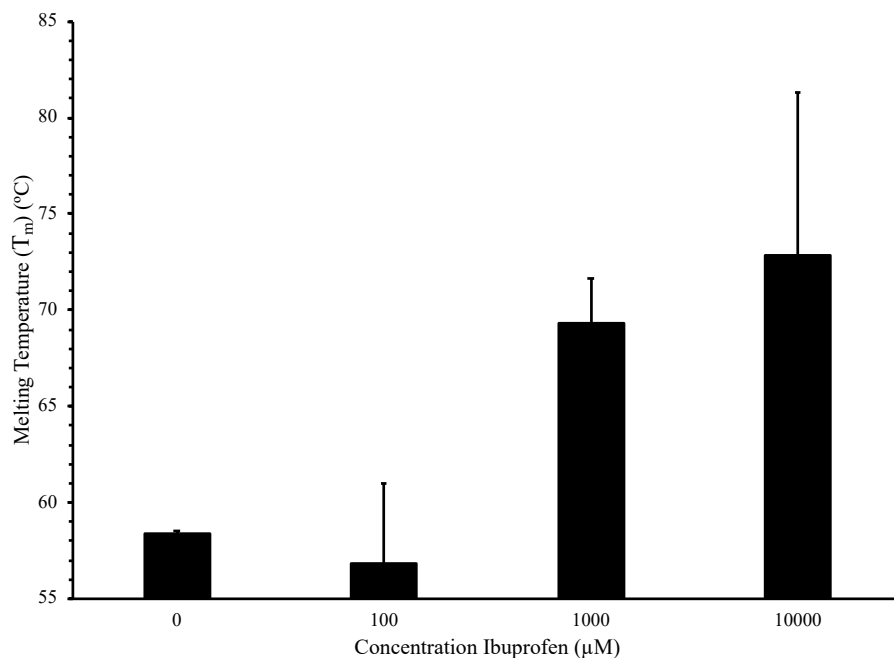


Figure 38. Average T_m of 10.0 μM BSA in the presence of varying concentrations of the drug ibuprofen. Ibuprofen is a high binder, with % protein binding around 90-99% bound. Averages of triplicate experiments. Standard deviations shown as error bars.

Conclusions

Based on the data presented here, we have initiated benchmarking of two methods that are likely to be effectively used to determine the difference between low and high protein binders. The fluorescence polarization (FP) method outlined above provides a more quantitative platform to do this, by using a competition assay between either fluorescein (the cheaper option, \$32.50 for 10 grams from Sigma Aldrich¹¹⁷) or BODIPY-LPA (a more expensive choice, \$203.41 for 1 milligram from Avanti Polar Lipids¹²⁰). If a significant decrease in mP is seen in competition with the report fluorophore, then the competitor is a higher binder to the protein. If a slight decrease or increase in mP is seen, then the competitor is a lower binder to the protein. The

normalization and scaling used in this method allow us to estimate % binding. This method also requires the purchase of a fluorophore, in addition to the BSA and compounds, which the NanoDSF method does not require. This method is also more time consuming than the NanoDSF method, as this technique requires parametrization for each fluorophore to be determined before the competition assays occur. Each assay takes more time than the NanoDSF method as well. The FP method requires 4.5 hours for each set of triplicate experiments, with a total of 13.5 hours for all three triplicate experiments required per compound. This method also requires more volumes of reagents be used per assay (14.4 mL used in the assay plate for all three triplicate experiments, plus the amount required to prepare all the solutions).

The NanoDSF method outlined above is a more qualitative way to determine a difference between high and low protein binders. If the T_m does not change as concentration of the compound increases, then the compound can be considered a low binder. If the T_m increases as concentration of the compound increases, then the compound can be considered a high binder. This method cannot determine how much of the compound is bound, however, which could be a drawback if that is the goal of the experiment. This method also does not require any additional costs other than those to purchase the BSA and compounds being tested. This method is quick (all three triplicate experiments can be performed in 4.5 hours, as the measurements themselves only take approximately 4 minutes) and easy to do, which is a benefit when examining many compounds at once. This method also only uses very little volume per assay (each capillary used in the Tycho uses only $\sim 10 \mu\text{L}$, so less than 1 mL is required for all three triplicate assays). To the best of our knowledge, this method has not been used to determine protein binding in this way before.

Chapter 4

Conclusions and Future Approaches

Summary of Major Points from Computational Approaches

In chapter 2, a discussion of the development of QSAR models predicting HIA, Caco-2 permeability, and protein binding was provided. These models were developed to help us make informed decisions about the suitability of potential compounds for subsequent ligand/inhibitor discovery efforts using ADMET properties (specifically absorption and distribution). These two ADMET properties were chosen because they are two of the initial processes an oral drug must progress through after ingestion.

Five models using different descriptor types for each HIA, Caco-2 permeability, and protein binding were developed with the goal to determine which descriptor type produced the predictions closest to reported experimental data. The development and analysis of these models was examined throughout chapter 2, along with a detailed examination of compounds commonly mispredicted within all 5 models for HIA, Caco-2 permeability, and protein binding. Although these models were successfully developed, they have not yet been used to make predictions on pipeline compounds with unknown ADMET properties. Such application awaits future students and or collaborative efforts. Instead of diverting effort to this endeavor, a decision was made to focus on experimental efforts that could be used to help validate these computational tools.

Model Selection and Future Application

The results of the computational work discussed in chapter 2 show that there was not a clear difference in model performance when comparing the models generated with different descriptor types and solvation approaches. The recommendation from this work is that future

models should be selected for future datasets depending on the size of the dataset to be examined. If the target dataset is smaller in size (less than 100 compounds), the i3D-Born descriptor model should be used. These models performed slightly better than all the other descriptor model types and better mimic physiological conditions, due to the use of Born solvation. However, these models require a conformational search to identify the lowest energy conformation of each structure. These searches take computational time, which is why if the dataset is larger in size (more than 100 compounds), the 2D descriptor model will be used. These models performed similarly to the i3D-Born descriptor type models, but do not require a conformational search because these descriptors do not depend on the three-dimensional space.

Examination of Mispredicted Compounds

Only one compound was commonly mispredicted within all 5 models for protein binding, so a histogram representing chemical diversity was not appropriate in this case. However, histograms comparing the mispredicted compounds in the HIA and Caco-2 models showed that the datasets were too diverse to uncover a trend in the structures of those compounds predicted to be in the incorrect category. The HIA mispredicted compounds histogram in Figure 14 had the majority of the pairwise similarities within the 20-40% bins, with all of the pairwise similarities falling between 20-70%. Only 2 of the 21 pairwise similarities fell within the 60-70% bin. This is not enough of a similarity to discern a common sub-structure within these mispredicted compounds to identify “bad actors”. The Caco-2 mispredicted compounds histogram in Figure 15 had the majority of the pairwise similarities within the 20-50% bins, with all of the pairwise similarities falling between 10-70%. Only 4 of the 104 pairwise similarities fell within the 60-70% bin. Again, this was too small of a number to discern a common sub-structure within these

mispredicted compounds. These data indicate that the dataset was too diverse to obtain an estimate of a commonality between the mispredicted compounds to find a way to filter out potentially troubling compounds in the future.

Summary of Major Points from Experimental Approaches

In chapter 3, a discussion of method development for protein binding experiments was presented. Protein binding experiments were conducted to help validate the protein binding models, described in chapter 2, and to make predictions of drug distribution as the overall goal to help prioritize ligand discovery efforts. Protein binding (PB) was the focus of these experimental methods, specifically binding to serum albumin. These protein binding experiments were conducted using both fluorescence polarization (FP) and nano differential scanning fluorimetry (NanoDSF) techniques. Both methods of protein binding assays were successfully developed and allowed determination of low and high serum albumin binding small molecules.

Fluorescence Polarization Experiments

The results of the fluorescence polarization experiments, detailed in chapter 3, indicate that a working method to determine protein binding was developed. Competition assays using either fluorescein or BODIPY-LPA as the fluorophore were both explored. This method can distinguish a difference between high and low protein binders in a more quantitative manner by examining the decrease in mP of the fluorophore. As the fluorophore is competed off the BSA through the dose-dependent increase in small molecule addition, the mP decreases. Because all of our data is on the same scale and normalized to that of the fluorophore alone, the decrease in mP can determine % protein binding. A large decrease in mP indicates a high protein binder and

a small decrease (or increase depending on the fluorophore being used) in mP indicates a low protein binder. In our hands, the use of fluorescein as the reporter fluorophore was more easily competed off due to its lower binding affinity to BSA as compared to BODIPY-LPA.

NanoDSF Experiments

The results of the NanoDSF experiments indicate that a functional, fast, low cost assay to examine small molecule binding to serum albumin was developed. Herein we have described a working method to distinguish between high and low protein binders based on a change in T_m of BSA in the presence of varying concentrations of compounds. This method is a qualitative way to determine protein binding, as there is no empirical way to determine how much of the small molecule is bound to BSA; there is only either no observable change in T_m for low binders or a change in T_m is observed for high binders. This qualitative method is fast, using only compound of interest in a label-free way, and therefore inexpensive. This method can be used in the future as a quick-pass way to determine compounds as high or low protein binders, as a way to filter out if a compound is worth pursuing in the fluorescence polarization experiments to determine a more quantitative number of protein binding. This approach, using the Tycho NT.6, is novel.

Recommendations for Future Experiments

It is recommended for future researchers to first use the quick-pass, inexpensive NanoDSF methodology to determine high vs. low protein binders. The compounds that are identified as low binders, based on no observed change in T_m , are then proposed to be examined in the fluorescence polarization experiments. This recommendation is based on the protein binding cutoffs outlined in chapter 2, where anything above 85% (high binding) negatively

impacts the drug's distribution. The optimal range for protein binding is 80-85%, so using the FP methods to quantitatively determine the protein binding of compounds suggested to be low binders in the NanoDSF methods will determine whether the compound is a true low binder (less than 80%) or if the compound fits within that 80-85% clinically relevant range. It is suggested that these FP experiments be performed using fluorescein as the fluorophore to be competed off the BSA.

Future Approaches

Additional Computational Work to Expand Established Methods

The next step, now that the models are built and the experimental methods are established, is to use the computational models to make predictions on compounds of interest. Compounds of interest to our research group include autotaxin (ATX) and SK inhibitors, LPA and S1P GPCR antagonists/agonists, as well as potential ligands for orphan GPCR, for example GPR88 and GPR52.

In particular, a major thrust in our research group has begun to focus on the deorphanization of several GPCRs, specifically GPR88 and GPR52, which are both thought to be associated with psychiatric disorders. GPR88 has been associated with susceptibility to both bipolar disorder and schizophrenia.¹³² GPR52 knockout mice showed psychosis-related behavior, which suggested that GPR52 may prove to be a therapeutic target for psychiatric disorders.¹³³

The initial compounds proposed for analysis via predictions of HIA, Caco-2 permeability, and protein binding using the QSAR models described in chapter 2 are shown in Figure 39. The ATX inhibitors proposed have no known pharmacokinetic (PK) data, namely Ragle 74¹³⁴ (a potent ATX inhibitor synthesized in our lab) and Banerjee 3b¹³⁵ (a potent ATX

inhibitor synthesized by collaborators at UTHSC). The LPA GPCR antagonists and agonists proposed were identified either by our lab or collaborators at UTHSC, namely H2L5186303¹³⁶, ¹³⁷ (LPA₂ antagonist identified by our lab), NSC161613¹³⁶ (LPA₃ antagonist identified by our lab), and GRI977143^{138, 139} (LPA₂ agonist identified by collaborators at UTHSC). The orphan GPCR potential ligands suggested are 2OMPP¹⁴⁰⁻¹⁴² (a synthetic GPR88 agonist) and 4u¹⁴³ (a synthetic GPR52 agonist).

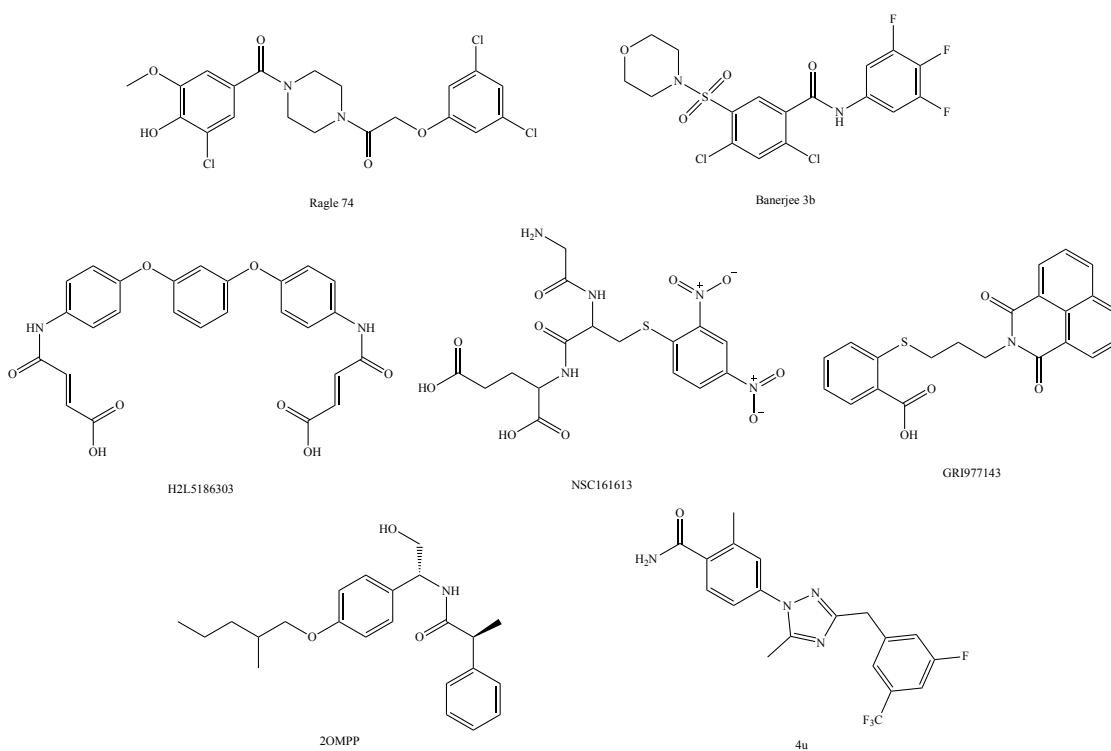


Figure 39. Structures of seven initial targets. Top two are ATX inhibitors, middle three are LPA GPCR ligands, and bottom two are orphan GPCR ligands.

Predictions of all three ADMET properties described herein will be completed using workflows described for these 7 initial potential therapeutics. Once complete experimental evaluation of protein binding will be verified using workflows described. Ideally, these

candidates would be predicted to be well absorbed for the HIA and Caco-2 permeability values and low protein binders. If these compounds fall within these categories, or if they fall within the moderately absorbed categories and low protein binding, then they will pass through to the experimental stage for analysis.

Additional Computational Methods to be Applied

Predicting Cytochrome P450 Interactions to Determine Potential Metabolism Reactions

Future computational efforts to help bolster the absorbance and distribution predictions described include analysis of potential metabolism by cytochrome (CYP) P450 enzymes. These enzymes are involved in phase I metabolism reactions in the liver and as such would bolster the initial phases of orally administered drug interactions.¹³ These enzymes have previously been targeted to determine substrate specificity, to examine inhibitors of these enzymes, and to determine ligand promiscuity for these enzymes.¹⁴⁴⁻¹⁴⁷ There are currently 212 protein data bank (PDB) entries in the RCSB protein data bank for human cytochrome P450 enzymes.¹⁴⁸ There are 57 isoforms of CYP450 enzymes found in humans,¹⁴⁹ but only 9 are responsible for common phase I metabolism reactions (CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, and CYP3A4).¹⁵⁰ There is a free tool available for download, developed in 2018, that predicts whether certain compounds interact with these CYP isoforms to produce metabolism reactants, called CypReact.⁶⁴ This tool can be used by future researchers to predict potentially dangerous metabolites of target compounds.

Additional Experimental Work Utilizing Established Methods

Experimental Determination of Protein Binding

Compounds that are successfully passed through the computational filters, will first be examined utilizing the qualitative NanoDSF methods outlined in chapter 3. This quick-pass experiment will be used to confirm whether the protein binding model predictions were correct (a high or low binder). If the NanoDSF experiments confirm that the model predictions were accurate, then fluorescence polarization experiments will be performed to quantify protein binding and further validate computational predictions and initial qualitative experimental results. If the NanoDSF control experiments show that there is some interference at those wavelengths (350nm/330nm), then fluorescence polarization experiments will be conducted directly to confirm computational predictions. As additional projects mature in the lab and with collaborators, additional compounds resulting from these efforts will be analyzed through the combined computational and experimental procedures outlined.

Additional Experimental Work

Method Development & Experimental Determination of Caco-2 Permeability

Further experiments to help filter potential therapeutics to be synthesized in the lab are Caco-2 permeability assays. These assays can help validate additional model predictions and narrow down compounds to be synthesized or to be further analyzed. Caco-2 permeability assays have been used previously to model the intestinal barrier, since the discovery that the cell-line exhibited similar characteristics to the human intestinal barrier¹⁵¹ and the first methods were optimized in the 1980s.¹⁵² This cell line can be easily purchased from the American Type Culture Collection (ATCC).¹⁵³ Typically, these cell lines are cultured for 21 days after seeding as a

monolayer on membrane inserts to reach confluence, after which the cell culture medium is removed from the apical (top) & basolateral (bottom) chambers of the apparatus.¹⁵⁴ This medium is replaced with the Permeability Assay Buffer (PAB) and the compounds evaluated are added to the apical side of the chamber, where the concentrations of each are assayed at varying timepoints to determine how much compound has permeated the membrane.¹⁵⁴ Such experiments require analytical methodology to be optimized for the quantitative analysis of each target compound. As such they would only be done for compounds passing all previous computational and experimental milestones. These methods will allow future efforts to help prioritize compounds for synthesis and further analysis as potential therapeutics.

Method Development and Experimental Determination of Metabolism

Significant amounts of phase I and II metabolism reactions occur in the liver, which is why researchers usually attempt to make determinations using liver microsomes. Human liver microsomes from 50 pooled donors are available from ThermoFisher Scientific and can be used for this purpose.¹⁵⁵ Methods outlined below utilizing liver microsomes to determine metabolites are based on methods outlined by Jeong, *et al.* in 2017.¹⁵⁶ Compounds would be incubated with liver microsomes (0.5 mg/mL for human microsomes) as well as reduced nicotinamide adenine dinucleotide phosphate (NADP⁻) generation cocktail consisting of: nicotinamide adenine dinucleotide phosphate (NADP⁺) at 1 mM, glucose-6-phosphate at 5 mM, glucose-6-phosphate dehydrogenase at 1 unit/mL, and magnesium chloride (MgCl₂) at 5 mM in sodium phosphate buffer (0.2 M, pH 7.4) so a total of 1 mL is reached. This mixture can be incubated in a shaking water bath & the enzymatic reaction is begun by the addition of NADP⁺. At varying timepoints, aliquots can be removed, prepped for analysis by mass spectrometry, or other methods as

appropriate to determine loss of parent compound as a first pass. The hepatic intrinsic clearance would then be calculated and compared to known reagents using tools developed in 1997 by Iwatsubo *et al.*¹⁵⁷

References

1. Anonymous. The Drug Development Process. <https://www.fda.gov/ForPatients/Approvals/Drugs/default.htm> (accessed Mar 20, 2020).
2. Yang, H.; Sun, L.; Li, W.; Liu, G.; Tang, Y. In Silico Prediction of Chemical Toxicity for Drug Design Using Machine Learning Methods and Structural Alerts. *Frontiers in chemistry* **2018**, *6*, 30.
3. Ban, F.; Dalal, K.; Li, H.; LeBlanc, E.; Rennie, P. S.; Cherkasov, A. Best Practices of Computer-Aided Drug Discovery: Lessons Learned from the Development of a Preclinical Candidate for Prostate Cancer with a New Mechanism of Action. *Journal of Chemical Information and Modeling* **2017**, *57*, 1018-1028.
4. Gifford, E.; van de Waterbeemd, H. ADMET In Silico Modelling: Towards Prediction Paradise? *Nature Reviews Drug Discovery* **2003**, *2*, 192-204.
5. Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A. An Analysis of the Attrition of Drug Candidates from Four Major Pharmaceutical Companies. *Nature reviews. Drug discovery* **2015**, *14*, 475-486.
6. DiMasi, J. A.; Feldman, L.; Seckler, A.; Wilson, A. Trends in Risks Associated with New Drug Development: Success Rates for Investigational Drugs. *Clinical Pharmacology & Therapeutics* **2010**, *87*, 272-277.
7. Takebe, T.; Imai, R.; Ono, S. The Current Status of Drug Discovery and Development as Originated in United States Academia: The Influence of Industrial and Academic Collaboration on Drug Discovery and Development. *Clinical and Translational Science* **2018**, *11*, 597-606.
8. Roberts, S. A. High-Throughput Screening Approaches for Investigating Drug Metabolism and Pharmacokinetics. *Xenobiotica* **2001**, *31*, 557-589.
9. Chohan, K.; Paine, S.; Waters, N. Advancements in Predictive In Silico Models for ADME. *Current Chemical Biology* **2008**, *2*, 215-228.
10. Sanders, J. M.; Beshore, D. C.; Culberson, J. C.; Fells, J. I.; Imbriglio, J. E.; Gunaydin, H.; Haidle, A. M.; Labroli, M.; Mattioni, B. E.; Sciammetta, N.; Shipe, W. D.; Sheridan, R. P.; Suen, L. M.; Verras, A.; Walji, A.; Joshi, E. M.; Bueters, T. Informing the Selection of Screening Hit Series with in Silico Absorption, Distribution, Metabolism, Excretion, and Toxicity Profiles. *Journal of medicinal chemistry* **2017**, *60*, 6771-6780.
11. Smith, D. A. Evolution of ADME Science: Where Else Can Modeling and Simulation Contribute? *Molecular Pharmaceutics* **2013**, *10*, 1162-1170.

12. Silverman, R. B.; Holladay, M. W. Introduction. In *The Organic Chemistry of Drug Design and Drug Action*; Academic Press: 2014; pp 1-17.
13. Bennet, L. Z.; Mitchell, J. R.; Sheiner, L. B. Pharmacokinetics: The Dynamics of Drug Absorption, Distribution, and Elimination. In *Goodman and Gilman's The Pharmacological Basis of Therapeutics*; Goodman Gilman, A., Rall, T. W., Nies, A. S. and Taylor, P., Eds.; McGraw-Hill, Inc.: 1993.
14. Mitscher, L. A. Drug Design and Discovery: An Overview. In *Textbook of Drug Design and Discovery*; Krogscaard-Larsen, P., Liljefors, T. and Madsen, U., Eds.; Taylor & Francis: 2002; pp 1-34.
15. Smith, D. A.; Waterbeemd, H. v. d. Pharmacokinetics and Metabolism in Early Drug Discovery. *Current Opinion in Chemical Biology* **1999**, *3*, 373-378.
16. Cheng, K.; Li, C.; Uss, A. S. Prediction of Oral Drug Absorption in Humans - From Cultured Cell Lines and Experimental Animals. *Expert Opinion on Drug Metabolism & Toxicology* **2008**, *4*, 581-590.
17. Li, A. P. Screening for Human ADME/Tox Drug Properties in Drug Discovery. *Drug Discovery Today* **2001**, *6*, 357-366.
18. Lin, L.; Wong, H. Predicting Oral Drug Absorption: Mini Review on Physiologically-Based Pharmacokinetic Models. *Pharmaceutics* **2017**, *9*, 41.
19. Chohan, K.; Paine, S.; Waters, N. Advancements in Predictive In Silico Models for ADME. *Current Chemical Biology* **2008**, *2*, 215-228.
20. de Souza Neto, Lauro Ribeiro; Moreira-Filho, J. T.; Neves, B. J.; Maidana, Rocío Lucía Beatriz Riveros; Guimarães, A. C. R.; Furnham, N.; Andrade, C. H.; Silva, J., Floriano Paes. In Silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery. *Frontiers in Chemistry* **2020**, *8*, 93.
21. Atterwill, C. K.; Wing, M. G. In Vitro Preclinical Lead Optimisation Technologies (PLOTs) in Pharmaceutical Development. *Toxicology letters* **2002**, *127*, 143-151.
22. Kar, S.; Leszczynski, J. Recent Advances of Computational Modeling for Predicting Drug Metabolism: A Perspective. *Current Drug Metabolism* **2017**, *18*, 1106-1122.
23. Guengerich, F. P. Mechanisms of Drug Toxicity and Relevance to Pharmaceutical Development. *Drug Metabolism and Pharmacokinetics* **2011**, *26*, 3-14.
24. Fredriksson, R.; Lagerström, M. C.; Lundin, L.-G.; Schiöth, H. B. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Molecular Pharmacology* **2003**, *63*, 1256-1272.

25. Takeda, S.; Kadowaki, S.; Haga, T.; Takaesu, H.; Mitaku, S. Identification of G Protein-Coupled Receptor Genes from the Human Genome Sequence. *FEBS Letters* **2002**, *520*, 97-101.
26. Chien, E. Y. T.; Liu, W.; Zhao, O.; Katritch, V.; Han, G. W.; Hanson, M. A.; Shi, L.; Newman, A. H.; Javitch, J. A.; Cherezov, V.; Stevens, R. C. Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. *Science* **2010**, *330*, 1091-1095.
27. Yang, J.; Villar, V. A. M.; Armando, I.; Jose, P. A.; Zeng, C. G Protein-Coupled Receptor Kinases: Crucial Regulators of Blood Pressure. *Journal of the American Heart Association* **2016**, *5*, n/a.
28. Bar-Shavit, R.; Maoz, M.; Kancharla, A.; Nag, J. K.; Agranovich, D.; Grisaru-Granovsky, S.; Uziely, B. G Protein-Coupled Receptors in Cancer. *International Journal of Molecular Sciences* **2016**, *17*, 1320.
29. Takeda, S.; Kadowaki, S.; Haga, T.; Takaesu, H.; Mitaku, S. Identification of G Protein-Coupled Receptor Genes from the Human Genome Sequence. *FEBS Letters* **2002**, *520*, 97-101.
30. Armstrong, J. F.; Faccenda, E.; Harding, S. D.; Pawson, A. J.; Southan, C.; Sharman, J. L.; Campo, B.; Cavanagh, D. R.; Alexander, S. P. H.; Davenport, A. P.; Spedding, M.; Davies, J. A. The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: Extending Immunopharmacology Content and Introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. *Nucleic Acids Research* **2019**, *48*, D1006-D1021.
31. Anonymous. Orphan and Other 7TM Receptors.
<https://www.guidetopharmacology.org/GRAC/FamilyDisplayForward?familyId=115&familyType=GPCR> (accessed Apr 3, 2020).
32. Castagna, D.; Budd, D. C.; Macdonald, S. J. F.; Jamieson, C.; Watson, A. J. B. Development of Autotaxin Inhibitors: An Overview of the Patent and Primary Literature. *Journal of Medicinal Chemistry* **2016**, *59*, 5604-5621.
33. Anliker, B.; Chun, J. Cell Surface Receptors in Lysophospholipid Signaling. *Seminars in Cell and Developmental Biology* **2004**, *15*, 457-465.
34. Aikawa, S.; Hashimoto, T.; Kano, K.; Aoki, J. Lysophosphatidic Acid as a Lipid Mediator with Multiple Biological Actions. *Journal of biochemistry* **2015**, *157*, 81-89.
35. Desroy, N.; Housseman, C.; Bock, X.; Joncour, A.; Bienvenu, N.; Cherel, L.; Labequere, V.; Rondet, E.; Peixoto, C.; Grassot, J.; Picolet, O.; Annoot, D.; Triballeau, N.; Monjardet, A.; Wakselman, E.; Roncoroni, V.; Le Tallec, S.; Blanque, R.; Cottreaux, C.; Vandervoort, N.; Christophe, T.; Mollat, P.; Lamers, M.; Auberval, M.; Hrvacic, B.; Ralic, J.; Oste, L.; van der Aar, E.; Brys, R.; Heckmann, B. Discovery of 2-[[2-Ethyl-6-

- [4-[2-(3-hydroxyazetid-1-yl)-2-oxoethyl]piperazin-1-yl]-8-methylimidazo[1,2-a]pyridin-3-yl]methylamino]-4-(4-fluorophenyl)thiazole-5-carbonitrile (GLPG1690), a First-in-Class Autotaxin Inhibitor Undergoing Clinical Evaluation for the Treatment of Idiopathic Pulmonary Fibrosis. *Journal of Medicinal Chemistry* **2017**, *60*, 3580-3590.
36. Stein, A. J.; Bain, G.; Prodanovich, P.; Santini, A. M.; Darlington, J.; Stelzer, N. M. P.; Sidhu, R. S.; Schaub, J.; Goulet, L.; Lonergan, D.; Calderon, I.; Evans, J. F.; Hutchinson, J. H. Structural Basis for Inhibition of Human Autotaxin by Four Potent Compounds with Distinct Modes of Binding. *Molecular Pharmacology* **2015**, *88*, 982-992.
37. Staehle, W.; Schiemann, K.; Schultz, M. WO2010112116A1, 2010.
38. Fells, J. I.; Lee, S. C.; Fujiwara, Y.; Norman, D. D.; Lim, K. G.; Tsukahara, R.; Liu, J.; Patil, R.; Miller, D. D.; Kirby, R. J.; Nelson, S.; Seibel, W.; Papoian, R.; Parrill, A. L.; Baker, D. L.; Bittman, R.; Tigyi, G. Hits of a High-Throughput Screen Identify the Hydrophobic Pocket of Autotaxin/Lysophospholipase D As an Inhibitory Surface. *Molecular Pharmacology* **2013**, *84*, 415-424.
39. Norman, D. D.; Ibezim, A.; Scott, W. E.; White, S.; Parrill, A. L.; Baker, D. L. Autotaxin Inhibition: Development and Application of Computational Tools to Identify Site-Selective Lead Compounds. *Bioorganic & Medicinal Chemistry* **2013**, *21*, 5548-5560.
40. Albers, Harald M. H. G.; Hendrickx, L. J. D.; van Tol, Rob J. P.; Hausmann, J.; Perrakis, A.; Ovaa, H. Structure-Based Design of Novel Boronic Acid-Based Inhibitors of Autotaxin. *Journal of Medicinal Chemistry* **2011**, *54*, 4619-4626.
41. Binder, B. Y. K.; Williams, P. A.; Silva, E. A.; Leach, J. K. Lysophosphatidic Acid and Sphingosine-1-Phosphate: A Concise Review of Biological Function and Applications for Tissue Engineering. *Tissue Engineering Part B: Reviews* **2015**, *21*, 531-542.
42. Ignatov, A.; Lintzel, J.; Kreienkamp, H.; Chica Schaller, H. Sphingosine-1-Phosphate is a High-Affinity Ligand for the G Protein-Coupled Receptor GPR6 from Mouse and Induces Intracellular Ca²⁺ Release by Activating the Sphingosine-Kinase Pathway. *Biochemical and Biophysical Research Communications* **2003**, *311*, 329-336.
43. Ghosh, T. K.; Bian, J.; Gill, D. L. Intracellular Calcium Release Mediated by Sphingosine Derivatives Generated in Cells. *Science* **1990**, *248*, 1653-1656.
44. Kim, M. Y.; Liang, G. H.; Kim, J. A.; Kim, Y. J.; Oh, S.; Suh, S. H. Sphingosine-1-Phosphate Activates BKCa Channels Independently of G Protein-Coupled Receptor in Human Endothelial Cells. *American Journal of Physiology - Cell Physiology* **2006**, *290*, 1000-1008.
45. Olivera, A.; Kohama, T.; Edsall, L.; Nava, V.; Cuvillier, O.; Poulton, S.; Spiegel, S. Sphingosine Kinase Expression Increases Intracellular Sphingosine-1-Phosphate and Promotes Cell Growth and Survival. *The Journal of Cell Biology* **1999**, *147*, 545-557.

46. Liu, H.; Toman, R. E.; Goparaju, S. K.; Maceyka, M.; Nava, V. E.; Sankala, H.; Payne, S. G.; Bektas, M.; Ishii, I.; Chun, J.; Milstien, S.; Spiegel, S. Sphingosine Kinase Type 2 Is a Putative BH3-only Protein That Induces Apoptosis. *Journal of Biological Chemistry* **2003**, *278*, 40330-40336.
47. Hait, N. C.; Oskeritzian, C. A.; Paugh, S. W.; Milstien, S.; Spiegel, S. Sphingosine Kinases, Sphingosine 1-Phosphate, Apoptosis and Diseases. *BBA - Biomembranes* **2006**, *1758*, 2016-2026.
48. Sadahira, Y.; Ruan, F.; Hakomori, S.-I.; Igarashi, Y. Sphingosine 1-Phosphate, a Specific Endogenous Signaling Molecule Controlling Cell Motility and Tumor Cell Invasiveness. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 9686-9690.
49. Zhang, H.; Desai, N. N.; Olivera, A.; Seki, T.; Brooker, G.; Spiegel, S. Sphingosine-1-Phosphate, a Novel Lipid, Involved in Cellular Proliferation. *The Journal of Cell Biology* **1991**, *114*, 155-167.
50. Olivera, A.; Spiegel, S. Sphingosine-1-Phosphate as Second Messenger in Cell Proliferation Induced by PDGF and FCS Mitogens. *Nature* **1993**, *365*, 557-560.
51. Spiegel, S.; Vanek, P. G.; Cuvillier, O.; Gutkind, J. S.; Pirianov, G.; Coso, O. A.; Kleuser, B. Suppression of Ceramide-Mediated Programmed Cell Death by Sphingosine-1-Phosphate. *Nature* **1996**, *381*, 800-803.
52. Sarkar, S.; Maceyka, M.; Hait, N. C.; Paugh, S. W.; Sankala, H.; Milstien, S.; Spiegel, S. Sphingosine Kinase 1 is Required for Migration, Proliferation and Survival of MCF-7 Human Breast Cancer Cells. *FEBS Letters* **2005**, *579*, 5313-5317.
53. Paugh, S. W.; Paugh, B. S.; Rahmani, M.; Kapitonov, D.; Almenara, J. A.; Kordula, T.; Milstien, S.; Adams, J. K.; Zipkin, R. E.; Grant, S.; Spiegel, S. A Selective Sphingosine Kinase 1 Inhibitor Integrates Multiple Molecular Therapeutic Targets in Human Leukemia. *Blood* **2008**, *112*, 1382-1391.
54. Lim, K. G.; Tonelli, F.; Li, Z.; Lu, X.; Bittman, R.; Pyne, S.; Pyne, N. J. FTY720 Analogues as Sphingosine Kinase 1 Inhibitors: Enzyme Inhibition Kinetics, Allosterism, Proteasomal Degradation, and Actin Rearrangement in MCF-7 Breast Cancer Cells. *The Journal of Biological Chemistry* **2011**, *286*, 18633-18640.
55. Jameson, D.; Croney, J. Fluorescence Polarization: Past, Present and Future. *Combinatorial Chemistry & High Throughput Screening* **2003**, *6*, 167-176.
56. Mohamadi, M.; Tschammer, N.; Breitsprecher, D. Quick Protein Binding Analysis by Label-Free Thermal Shift Analysis on the Tycho NT.6. *Technical Note*.

57. Lagorce, D.; Sperandio, O.; Galons, H.; Miteva, M. A.; Villoutreix, B. O. FAF-Drugs2: Free ADME/tox Filtering Tool to Assist Drug Discovery and Chemical Biology Projects. *BMC Bioinformatics* **2008**, *9*, 396.
58. Cheng, F.; Li, W.; Zhou, Y.; Shen, J.; Wu, Z.; Liu, G.; Lee, P. W.; Tang, Y. admetSAR: A Comprehensive Source and Free Tool for Assessment of Chemical ADMET Properties. *Journal of Chemical Information and Modeling* **2012**, *52*, 3099-3105.
59. Cheng, F.; Li, W.; Zhou, Y.; Shen, J.; Wu, Z.; Liu, G.; Lee, P. W.; Tang, Y. Correction to “admetSAR: A Comprehensive Source and Free Tool for Assessment of Chemical ADMET Properties”. *Journal of Chemical Information and Modeling* **2019**, *59*, 4959.
60. He, Y.; Liew, C. Y.; Sharma, N.; Woo, S. K.; Chau, Y. T.; Yap, C. W. PaDEL-DDPredictor: Open-Source Software for PD-PK-T Prediction. *Journal of Computational Chemistry* **2013**, *34*, 604-610.
61. Cao, D.; Xu, Q.; Hu, Q.; Liang, Y. ChemoPy: Freely Available Python Package for Computational Biology and Chemoinformatics. *Bioinformatics (Oxford, England)* **2013**, *29*, 1092-1094.
62. Schyman, P.; Liu, R.; Desai, V.; Wallqvist, A. vNN Web Server for ADMET Predictions. *Frontiers in Pharmacology* **2017**, *8*, 889.
63. Dong, J.; Wang, N.; Yao, Z.; Zhang, L.; Cheng, Y.; Ouyang, D.; Lu, A.; Cao, D. ADMETlab: a Platform for Systematic ADMET Evaluation Based on a Comprehensively Collected ADMET Database. *Journal of Cheminformatics* **2018**, *10*, 1-11.
64. Tian, S.; Djoumbou-Feunang, Y.; Greiner, R.; Wishart, D. S. CypReact: A Software Tool for in Silico Reactant Prediction for Human Cytochrome P450 Enzymes. *Journal of Chemical Information and Modeling* **2018**, *58*, 1282-1291.
65. Anonymous. Simulations Plus Releases ADMET Predictor Version 9.5. *Manufacturing Close-Up* **2019**.
66. Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terflath, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *Journal of Medicinal Chemistry* **2014**, *57*, 4977-5010.
67. Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure. *Journal of Chemical Information and Computer Sciences* **1998**, *38*, 726-735.
68. Sambuy, Y.; De Angelis, I.; Ranaldi, G.; Scarino, M.; Stamatii, A.; Zucco, F. The Caco-2 Cell Line as a Model of the Intestinal Barrier: Influence of Cell and Culture-Related

- Factors on Caco-2 Cell Functional Characteristics. *Cell Biology and Toxicology* **2005**, *21*, 1-26.
69. van Breemen, R. B.; Li, Y. Caco-2 Cell Permeability Assays to Measure Drug Absorption. *Expert Opinion on Drug Metabolism & Toxicology* **2005**, *1*, 175-185.
70. Scheife, R. T. Protein Binding: What Does it Mean? *The Annals of Pharmacotherapy* **1989**, *23*, 27.
71. Maloney, P. P.; Muir, R. M.; Fujita, T.; Hansch, C. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178-180.
72. Hammett, L. P. Some Relations Between Reaction Rates and Equilibrium Constants. *Chemical Reviews* **1935**, *17*, 125-136.
73. Hansch, C.; Muir, R. M.; Fujita, T.; Maloney, P. P.; Geiger, F.; Streich, M. The Correlation of Biological Activity of Plant Growth Regulators and Chloromycetin Derivatives with Hammett Constants and Partition Coefficients. *Journal of the American Chemical Society* **1963**, *85*, 2817-2824.
74. Hansch, C.; Fujita, T. p- σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society* **1964**, *86*, 1616-1626.
75. Hansch, C. Quantitative Approach to Biochemical Structure-Activity Relationships. *Accounts of Chemical Research* **1969**, *2*, 232-239.
76. Leach, A. R. Quantitative Structure-Activity Relationships. In *Molecular Modelling--Principles and Applications*; Pearson Education Limited: pp 695-711.
77. Martin, Y. C. Hansch Analysis 50 Years On. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 435-442.
78. Lagorce, D.; Reynes, C.; Camproux, A.; Miteva, M. A.; Sperandio, O.; Villoutreix, B. O. In Silico ADME/Tox Predictions. In *ADMET for Medicinal Chemists*; John Wiley & Sons, Inc: Hoboken, NJ, USA, 2011; pp 29-124.
79. Roy, K.; Kar, S.; Das, R. N. Chapter 2 - Chemical Information and Descriptors. In *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*; Elsevier Inc: 2015; pp 47-80.
80. Leach, A. R. Partial Least Squares. In *Molecular Modelling: Principles and Applications*; Pearson Education Limited: 2001; pp 706-708.

81. Gao, H.; Lajiness, M. S.; Drie, J. V. Enhancement of Binary QSAR Analysis by a GA-Based Variable Selection Method. *Journal of Molecular Graphics and Modelling* **2002**, *20*, 259-268.
82. Gao, H.; Williams, C.; Labute, P.; Bajorath, J. Binary Quantitative Structure–Activity Relationship (QSAR) Analysis of Estrogen Receptor Ligands. *Journal of Chemical Information and Computer Sciences* **1999**, *39*, 164-168.
83. Hou, T.; Wang, J.; Zhang, W.; Xu, X. ADME Evaluation in Drug Discovery. 7. Prediction of Oral Absorption by Correlation and Classification. *Journal of Chemical Information and Modeling* **2007**, *47*, 208-218.
84. 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.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: a Major Update to the DrugBank Database for 2018. *Nucleic Acids Research* **2018**, *46*, D1074-D1082.
85. O'Hagan, S.; Kell, D. B. The Apparent Permeabilities of Caco-2 Cells to Marketed Drugs: Magnitude, and Independence from Both Biophysical Properties and Endogenite Similarities. *PeerJ* **2015**, *3*, e1405.
86. Fonger, G. C.; Hakkinen, P.; Jordan, S.; Publicker, S. The National Library of Medicine's (NLM) Hazardous Substances Data Bank (HSDB): Background, Recent Enhancements and Future Plans. *Toxicology* **2014**, *325*, 209-216.
87. Radchenko, E.; Dyabina, A.; Palyulin, V.; Zefirov, N. Prediction of Human Intestinal Absorption of Drug Compounds. *Russian Chemical Bulletin* **2016**, *65*, 576-580.
88. Yee, S. In Vitro Permeability Across Caco-2 Cells (Colonic) Can Predict In Vivo (Small Intestinal) Absorption in Man—Fact or Myth. *Pharmaceutical Research* **1997**, *14*, 763-766.
89. Scheife, R. T. Protein Binding: What Does it Mean? *The Annals of Pharmacotherapy* **1989**, *23*, 27.
90. Clark, R. D.; Daga, P. R. Building a Quantitative Structure-Property Relationship (QSPR) Model. *Methods in Molecular Biology (Clifton, N.J.)* **2019**, *1939*, 139-159.
91. González-Medina, M.; Prieto-Martínez, F.; Owen, J.; Medina-Franco, J. Consensus Diversity Plots: a Global Diversity Analysis of Chemical Libraries. *Journal of Cheminformatics* **2016**, *8*, 1-11.
92. Anonymous. Molecular Operating Environment (MOE). **2016**.

93. Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *Journal of Chemical Information and Computer Sciences* **2002**, *42*, 1273-1280.
94. Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *Journal of Cheminformatics* **2015**, *7*, 1-13.
95. Halgren, T. A. MMFF VI. MMFF94s Option for Energy Minimization Studies. *Journal of Computational Chemistry* **1999**, *20*, 720-729.
96. Topliss, J. G.; Costello, R. J. Chance Correlations in Structure-Activity Studies Using Multiple Regression Analysis. *Journal of Medicinal Chemistry* **1972**, *15*, 1066-1068.
97. Pearson, K. Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London* **1895**, *58*, 240-242.
98. Cramer, H. *Mathematical Methods of Statistics (PMS-9)*; Princeton University Press: United States, 1946.
99. Conover, W. J. *Practical nonparametric statistics*; United States, 1999.
100. Press, William H. (Mitverf.) *Numerical recipes. The art of scientific computing*; Cambridge University Press: Cambridge, 2007.
101. Tanis, E. A. *Probability and Statistical Inference*; Macmillan Publishing: New York, 1993.
102. Gronau, Q.; Gronau, Q.; Wagenmakers, E.; Wagenmakers, E. Limitations of Bayesian Leave-One-Out Cross-Validation for Model Selection. *Computational Brain & Behavior* **2019**, *2*, 1-11.
103. Silverman, R. B.; Holladay, M. W. Drug and Receptor Chirality. In *The Organic Chemistry of Drug Design and Drug Action*; Elsevier, Inc.: 2014; pp 140-145.
104. McConathy, J.; Owens, M. J. Stereochemistry in Drug Action. *Primary Care Companion to the Journal of Clinical Psychiatry* **2003**, *5*, 70-73.
105. Bohnert, T.; Gan, L. Plasma Protein Binding: From Discovery to Development. *Journal of Pharmaceutical Sciences* **2013**, *102*, 2953-2994.
106. Anonymous. Rapid Equilibrium Dialysis (RED) Device Single-Use Plate with Inserts, 8K MWCO. <https://www.thermofisher.com/order/catalog/product/90006#/90006> (accessed Apr 3, 2020).
107. Uri, A.; Nonga, O. E. What is the Current Value of Fluorescence Polarization Assays in Small Molecule Screening? *Expert Opinion on Drug Discovery* **2020**, *15*, 131-133.

108. Magnusson, A. O.; Szekrenyi, A.; Joosten, H.; Finnigan, J.; Charnock, S.; Fessner, W. nanoDSF as Screening Tool for Enzyme Libraries and Biotechnology Development. *The FEBS Journal* **2019**, *286*, 184-204.
109. Roehrl, M. H. A.; Wang, J. Y.; Wagner, G. A General Framework for Development and Data Analysis of Competitive High-Throughput Screens for Small-Molecule Inhibitors of Protein–Protein Interactions by Fluorescence Polarization. *Biochemistry* **2004**, *43*, 16056-16066.
110. Moerke, N. J. Fluorescence Polarization (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding. *Current Protocols in Chemical Biology* **2009**, *1*, 1-15.
111. Anonymous. FlexStation 3 Multi-Mode Microplate Reader: A Five-Mode Microplate Reader with Integrated Fluid Transfer. *Data Sheet* **2016**.
112. Barbero, N.; Barni, E.; Barolo, C.; Quagliotto, P.; Viscardi, G.; Napione, L.; Pavan, S.; Bussolino, F. A Study of the Interaction Between Fluorescein Sodium Salt and Bovine Serum Albumin by Steady-State Fluorescence. *Dyes and Pigments* **2009**, *80*, 307-313.
113. Tigyi, G.; Miledi, R. Lysophosphatidates Bound to Serum Albumin Activate Membrane Currents in Xenopus Oocytes and Neurite Retraction in PC12 Pheochromocytoma Cells. *Journal of Biological Chemistry* **1992**, *267*, 21360.
114. Saunders, L. P.; Cao, W.; Chang, W. C.; Albright, R. A.; Braddock, D. T.; De La Cruz, Enrique, M. Kinetic Analysis of Autotaxin Reveals Substrate-specific Catalytic Pathways and a Mechanism for Lysophosphatidic Acid Distribution. *The Journal of Biological Chemistry* **2011**, *286*, 30130-30141.
115. Penniston, J. T. Fluorescence Polarization Measurement of Binding of Fluorescein to Albumin. *Experimental Eye Research* **1982**, *34*, 435-443.
116. Brubaker, R. F.; Penniston, J. T.; Grotte, D. A.; Nagataki, S. Measurement of Fluorescein Binding in Human Plasma Using Fluorescence Polarization. *Archives of Ophthalmology* **1982**, *100*, 625-630.
117. Anonymous. Fluorescein Sodium Salt.
<https://www.sigmaaldrich.com/catalog/substance/fluoresceinsodiumsalt3762751847811?lang=en®ion=US> (accessed Apr 3, 2020).
118. Anonymous. Fluorescence Polarization (FP)--Note 1.4.
<https://www.thermofisher.com/us/en/home/references/molecular-probes-the-handbook/technical-notes-and-product-highlights/fluorescence-polarization-fp.html> (accessed Nov 5, 2019).

119. Chagas-Lima, A.; Pereira, M.; Fampa, P.; Lima, M.; Kluck, G.; Atella, G. Bioactive lipids Regulate Trypanosoma Cruzi Development. *Parasitology Research* **2019**, *118*, 2609-2619.
120. Anonymous. TopFluor Lyso PA. <https://avantilipids.com/product/810280> (accessed Apr 3, 2020).
121. Chattopadhyay, G.; Varadarajan, R. Facile Measurement of Protein Stability and Folding Kinetics Using a Nano Differential Scanning Fluorimeter. *Protein Science* **2019**, *28*, 1127-1134.
122. McEvoy, G. K. In *American Hospital Formulary Service--Drug Information 2003*. American Society of Health-System Pharmacists, Inc.: Bethesda, MD, 2003; pp 1743.
123. Anonymous. In *Drug Information 2018*. American Society of Health-System Pharmacists: Bethesda, MD, 2018; pp 2888.
124. Bushra, R.; Aslam, N. An Overview of Clinical Pharmacology of Ibuprofen. *Oman Medical Journal* **2010**, *25*, 155-161.
125. Anonymous. Naprelan Drug Label.
126. Anonymous. Naprosyn Drug Label.
127. 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.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: a Major Update to the DrugBank Database for 2018. *Nucleic Acids Research* **2018**, *46*, D1074-D1082.
128. Mathias, U.; Jung, M. Determination of Drug–Serum Protein Interactions via Fluorescence Polarization Measurements. *Analytical and Bioanalytical Chemistry* **2007**, *388*, 1147-1156.
129. Zhang, C.; Nordeen, S. K.; Shapiro, D. J. Fluorescence Anisotropy Microplate Assay to Investigate the Interaction of Full-Length Steroid Receptor Coactivator-1a with Steroid Receptors. *Methods in Molecular Biology (Clifton, N.J.)* **2013**, *977*, 339-351.
130. Anonymous. Protein Fluorescence. In *Principles of Fluorescence Spectroscopy*; Lakowicz, J. R., Ed.; Springer: Boston, MA, 2006.
131. Anonymous. Protein Fluorescence. In *Principles of Fluorescence Spectroscopy*; Lakowicz, J. R., Ed.; Springer: Boston, MA, 2006.
132. Del Zompo, M.; Deleuze, J.; Chillotti, C.; Cousin, E.; Niehaus, D.; Ebstein, R. P.; Arda, R.; Macé, S.; Warnich, L.; Mujahed, M.; Severino, G.; Dib, C.; Jordaan, E.; Murad, I.;

- Soubigou, S.; Koen, L.; Bannoura, I.; Rocher, C.; Laurent, C.; Derock, M.; Faucon Biguet, N.; Mallet, J.; Meloni, R. Association Study in Three Different Populations Between the GPR88 Gene and Major Psychoses. *Molecular Genetics & Genomic Medicine* **2014**, *2*, 152-159.
133. Komatsu, H.; Maruyama, M.; Yao, S.; Shinohara, T.; Sakuma, K.; Imaichi, S.; Chikatsu, T.; Kuniyeda, K.; Siu, F. K.; Peng, L. S.; Zhuo, K.; Mun, L. S.; Han, T. M.; Matsumoto, Y.; Hashimoto, T.; Miyajima, N.; Itoh, Y.; Ogi, K.; Habata, Y.; Mori, M. Anatomical Transcriptome of G Protein-Coupled Receptors Leads to the Identification of a Novel Therapeutic Candidate GPR52 for Psychiatric Disorders. *PLoS one* **2014**, *9*, e90134.
134. Ragle, L. E. Computational and Experimental Tools to Explore Autotaxin Inhibition by Novel Small Molecules, ProQuest Dissertations Publishing, 2016.
135. Banerjee, S.; Norman, D. D.; Lee, S. C.; Parrill, A. L.; Pham, T. C. T.; Baker, D. L.; Tigyi, G. J.; Miller, D. D. Highly Potent Non-Carboxylic Acid Autotaxin Inhibitors Reduce Melanoma Metastasis and Chemotherapeutic Resistance of Breast Cancer Stem Cells. *Journal of Medicinal Chemistry* **2017**, *60*, 1309-1324.
136. Fells, J. I.; Tsukahara, R.; Fujiwara, Y.; Liu, J.; Perygin, D. H.; Osborne, D. A.; Tigyi, G.; Parrill, A. L. Identification of Non-Lipid LPA3 Antagonists by Virtual Screening. *Bioorganic & Medicinal Chemistry* **2008**, *16*, 6207-6217.
137. Fells, J. I.; Tsukahara, R.; Liu, J.; Tigyi, G.; Parrill, A. L. Structure-Based Drug Design Identifies Novel LPA3 Antagonists. *Bioorganic & Medicinal Chemistry* **2009**, *17*, 7457-7464.
138. Kiss, G. N.; Fells, J. I.; Gupte, R.; Lee, S.; Liu, J.; Nusser, N.; Lim, K. G.; Ray, R. M.; Lin, F.; Parrill, A. L.; Sümegi, B.; Miller, D. D.; Tigyi, G. Virtual Screening for LPA2-Specific Agonists Identifies a Nonlipid Compound with Antiapoptotic Actions. *Molecular Pharmacology* **2012**, *82*, 1162-1173.
139. Kiss, G. N.; Lee, S. C.; Fells, J. I.; Liu, J.; Valentine, W. J.; Fujiwara, Y.; Thompson, K. E.; Yates, C. R.; Sümegi, B.; Tigyi, G. Mitigation of Radiation Injury by Selective Stimulation of the LPA(2) Receptor. *Biochimica et Biophysica Acta* **2013**, *1831*, 117-125.
140. Dzierba, C. D.; Bi, Y.; Dasgupta, B.; Hartz, R. A.; Ahuja, V.; Cianchetta, G.; Kumi, G.; Dong, L.; Aleem, S.; Fink, C.; Garcia, Y.; Green, M.; Han, J.; Kwon, S.; Qiao, Y.; Wang, J.; Zhang, Y.; Liu, Y.; Zipp, G.; Liang, Z.; Burford, N.; Ferrante, M.; Bertekap, R.; Lewis, M.; Cacace, A.; Grace, J.; Wilson, A.; Nouraldeen, A.; Westphal, R.; Kimball, D.; Carson, K.; Bronson, J. J.; Macor, J. E. Design, Synthesis, and Evaluation of Phenylglycinols and Phenyl Amines as Agonists of GPR88. *Bioorganic & Medicinal Chemistry Letters* **2015**, *25*, 1448-1452.

141. Cianchetta, G.; Bronson, J. J.; Kwon, S.; Bi, Y.; Zhang, Y.; Dong, L.; Macor, J. E.; Carson, K.; Dzierba, C. D.; Green, M.; Wang, J.; Kimball, D.; Fink, C.; Zipp, G. CN102686575A, 2012.
142. Jin, C.; Decker, A. M.; Langston, T. L. Design, Synthesis and Pharmacological Evaluation of 4-Hydroxyphenylglycine and 4-Hydroxyphenylglycinol Derivatives as GPR88 Agonists. *Bioorganic & Medicinal Chemistry* **2017**, *25*, 805-812.
143. Tokumaru, K.; Ito, Y.; Nomura, I.; Nakahata, T.; Shimizu, Y.; Kurimoto, E.; Aoyama, K.; Aso, K. Design, Synthesis, and Pharmacological Evaluation of 4-Azoly-Benzamide Derivatives as Novel GPR52 Agonists. *Bioorganic & Medicinal Chemistry* **2017**, *25*, 3098-3115.
144. Cheng, F.; Li, W.; Liu, G.; Tang, Y. In Silico ADMET Prediction: Recent Advances, Current Challenges and Future Trends. *Current Topics in Medicinal Chemistry* **2013**, *13*, 1273-1289.
145. Mishra, N. K.; Agarwal, S.; Raghava, G. P. Prediction of Cytochrome P450 Isoform Responsible for Metabolizing a Drug Molecule. *BMC pharmacology* **2010**, *10*, 8.
146. Zientek, M.; Stoner, C.; Ayscue, R.; Klug-McLeod, J.; Jiang, Y.; West, M.; Collins, C.; Ekins, S. Integrated in Silico–in Vitro Strategy for Addressing Cytochrome P450 3A4 Time-Dependent Inhibition. *Chemical Research in Toxicology* **2010**, *23*, 664-676.
147. Nath, A.; Zientek, M. A.; Burke, B. J.; Jiang, Y.; Atkins, W. M. Quantifying and Predicting the Promiscuity and Isoform Specificity of Small-Molecule Cytochrome P450 Inhibitors. *Drug Metabolism and Disposition: the Biological Fate of Chemicals* **2010**, *38*, 2195-2203.
148. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28*, 235-242.
149. Tyzack, J. D.; Kirchmair, J. Computational Methods and Tools to Predict Cytochrome P450 Metabolism for Drug Discovery. *Chemical Biology & Drug Design* **2019**, *93*, 377-386.
150. Nebert, D. W.; Russell, D. W. Clinical Importance of the Cytochromes P450. *The Lancet* **2002**, *360*, 1155-1162.
151. Pinto, M.; Robine-Leon, S.; Appay, M.; Kedinger, M.; Triadou, N.; Dussaulx, E.; Lacroix, B.; Simon-Assman, P.; Haffen, K.; Fogh, J.; Zweibaum, A. Enterocyte-like Differentiation and Polarization of the Human Colon Carcinoma Cell Line Caco-2 in Culture. *Biology of the Cell* **1983**, *47*, 323-330.

152. Hidalgo, I. J.; Raub, T. J.; Borchardt, R. T. Characterization of the Human Colon Carcinoma Cell Line (Caco-2) as a Model System for Intestinal Epithelial Permeability. *Gastroenterology* **1989**, *96*, 736-749.
153. Anonymous. Caco-2 Product Information. <https://www.atcc.org/Products/All/HTB-37.aspx#generalinformation> (accessed Apr 3, 2020).
154. Press, B. Chapter 9--Optimization of the Caco-2 Permeability Assay to Screen Drug Compounds for Intestinal Absorption and Efflux. In *Permeability Barrier: Methods and Protocols. Methods in Molecular Biology*; Turksen, K., Ed.; Springer Science+Business Media, LLC: 2011; pp 139-154.
155. Anonymous. Human Microsomes, 50 Donors. <https://www.thermofisher.com/order/catalog/product/HMMCPL#/HMMCPL> (accessed Apr 3, 2020).
156. Jeong, H.; Kim, S.; Lee, J.; Park, J. Y.; Zhou, W.; Liu, X.; Kim, S. D.; Song, Y. S.; Jang, C.; Oh, S.; Choi, S.; Chang, M. Characterization of Phase I and Phase II Hepatic Metabolism and Reactive Intermediates of Larrea nitida Cav. and Its Lignan Compounds. *Phytotherapy Research* **2017**, *31*, 140-151.
157. Iwatsubo, T.; Suzuki, H.; Sugiyama, Y. Prediction of Species Differences (Rats, Dogs, Humans) in the In Vivo Metabolic Clearance of YM796 by the Liver from In Vitro Data. *Journal of Pharmacology and Experimental Therapeutics* **1997**, *283*, 462.

Appendix

Chapter 2 supplementary information. Boundary comparison figures for training and test set compounds used to determine the compounds within each dataset were ideal, the 19 remaining compound similarity histograms generated from similarity matrices used to ensure a broad chemical space was represented in the models (mentioned with Figure 6), and mispredicted compounds for each descriptor model type for both HIA and Caco-2 permeability models used to generate similarity histograms in Figures 10 & 11.

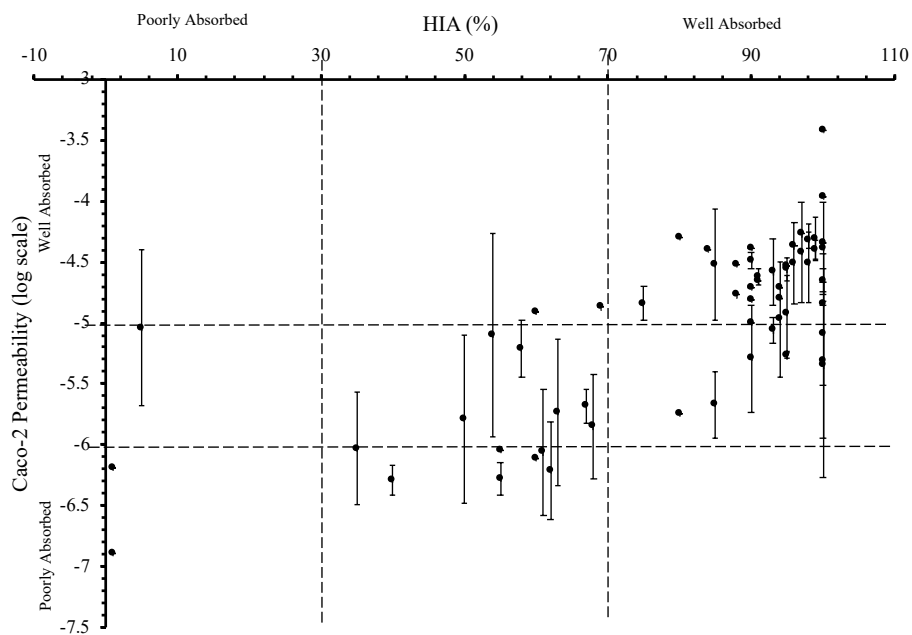


Figure A1. Boundary comparisons for training set compounds that determine absorption categories. The majority of compounds examined fall within the ideal diagonal.

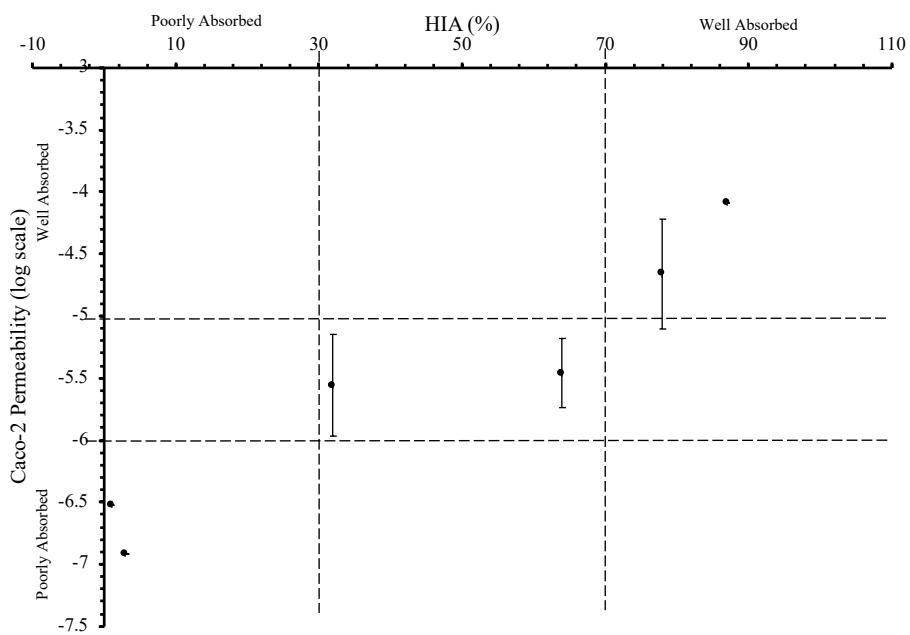


Figure A2. Boundary comparisons for test set compounds that determine absorption categories. All compounds fall within the ideal diagonal.

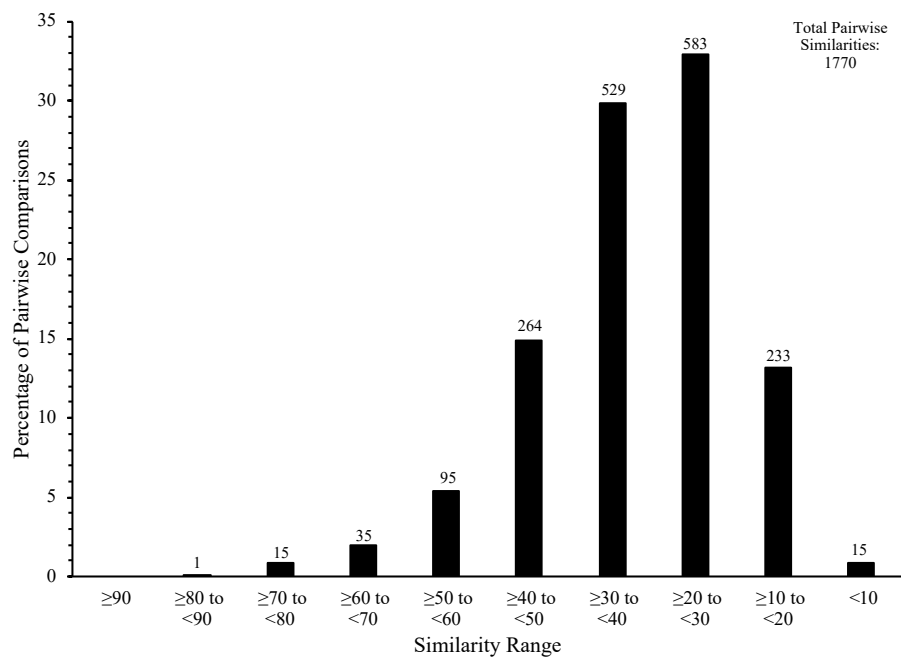


Figure A3. Similarity histogram for training set 1 compounds. Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of pairwise comparisons fall within the 10-50% similarity range.

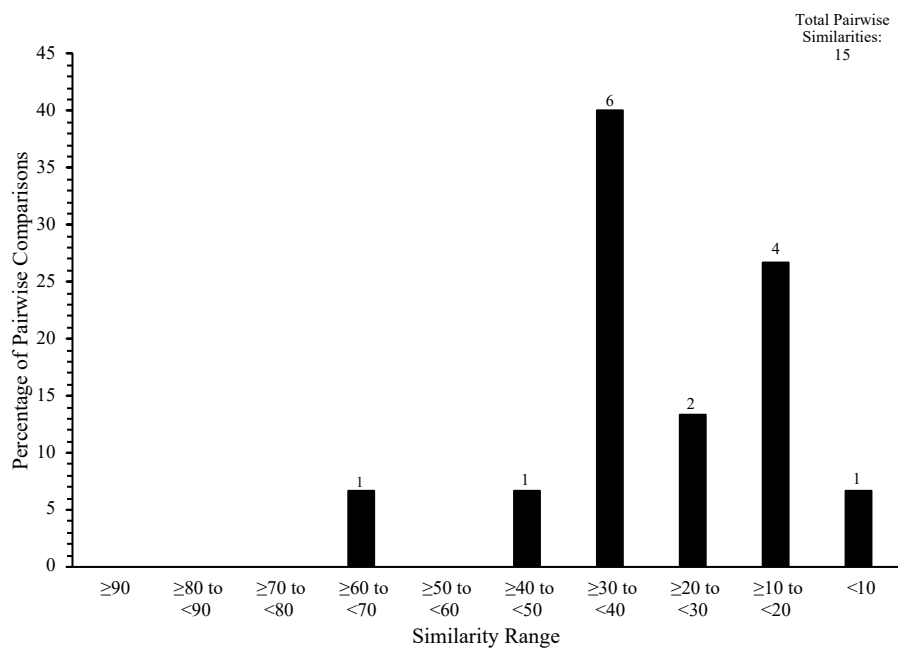


Figure A4. Similarity histogram for test set 1 compounds. Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. All of the pairwise comparisons fall within the 0-70% similarity range.

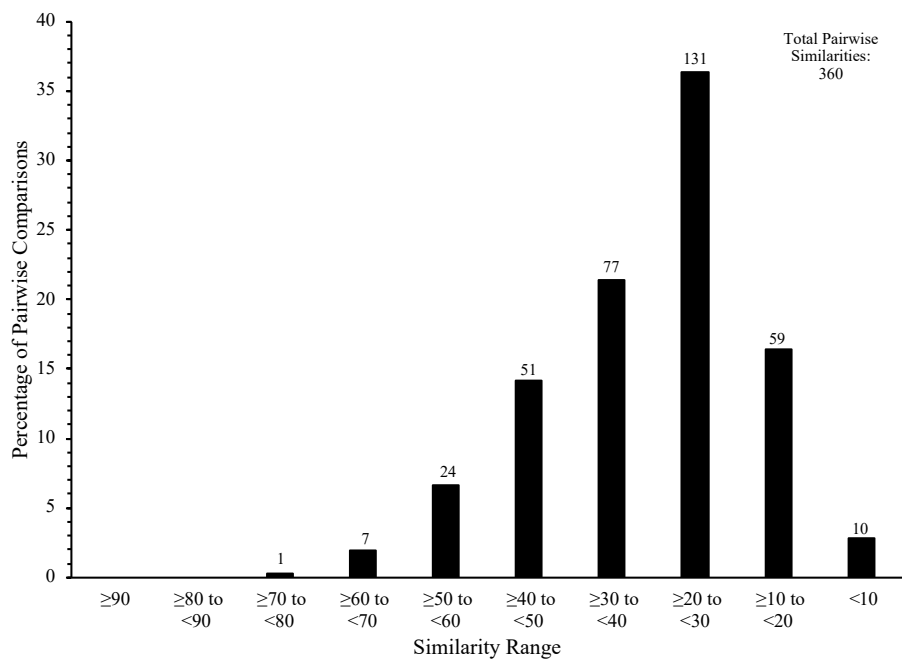


Figure A5. Similarity histogram for test set 1 vs. training set 1 compounds. Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of the pairwise comparisons fall within the 10-50% similarity range.

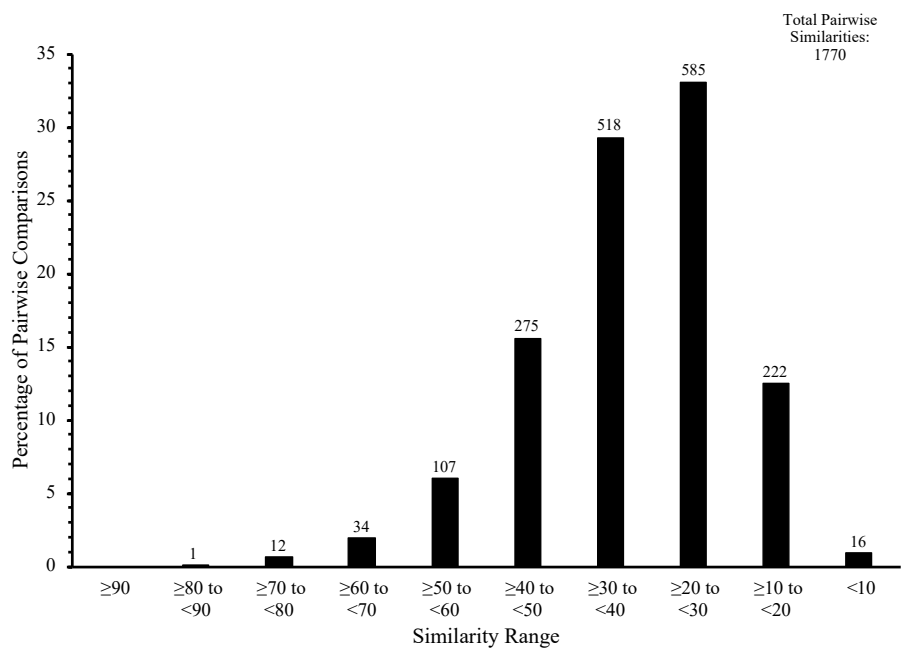


Figure A6. Similarity histogram for training set 2 compounds. Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of pairwise comparisons fall within the 10-50% similarity range.

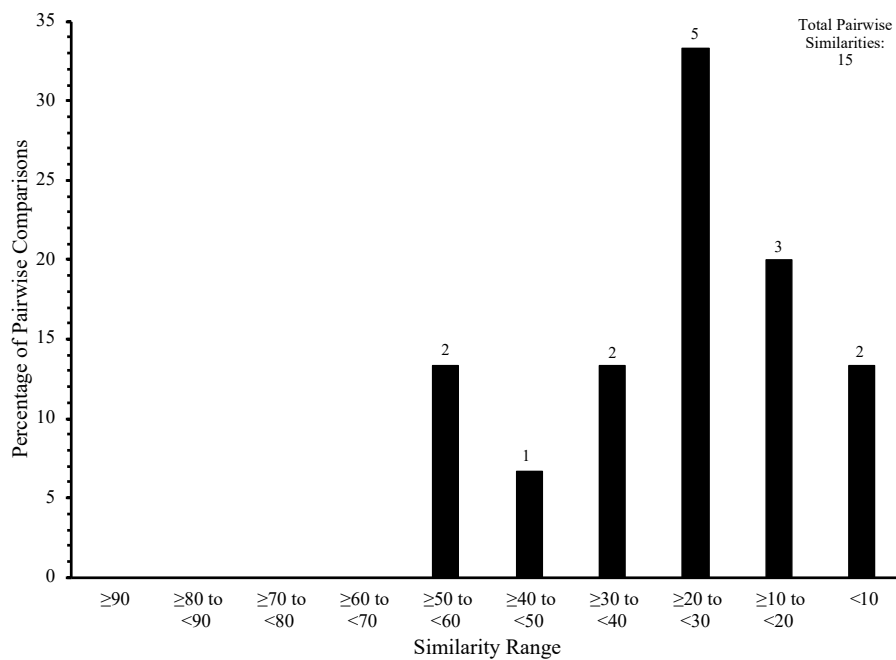


Figure A7. Similarity histogram for test set 2 compounds. Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. All of the pairwise comparisons fall within the 0-60% similarity range.

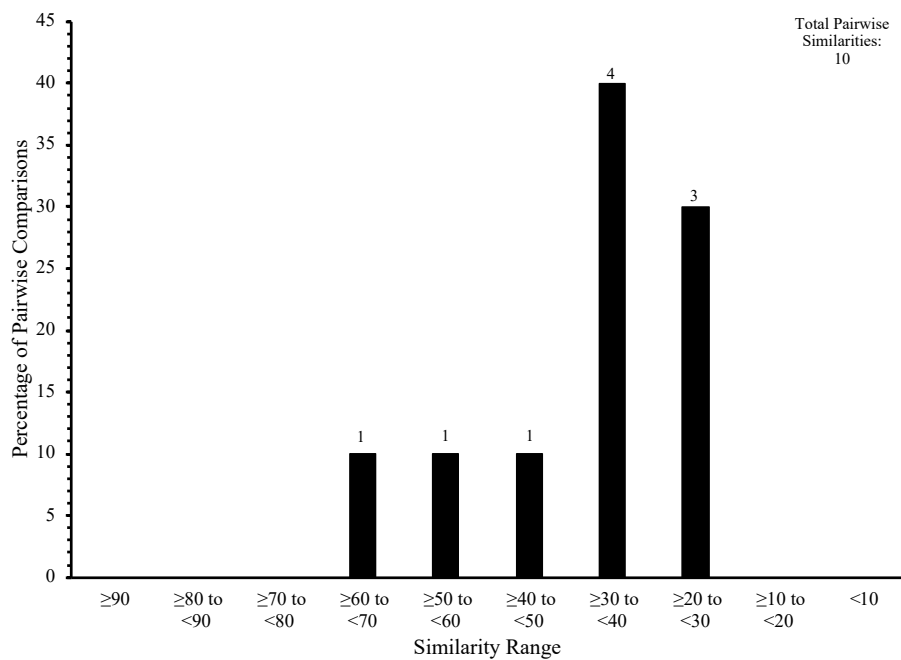


Figure A8. Similarity histogram for poorly absorbed compounds (both training and test set 1 & 2). Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. All the pairwise comparisons fall within the 20-70% similarity range.

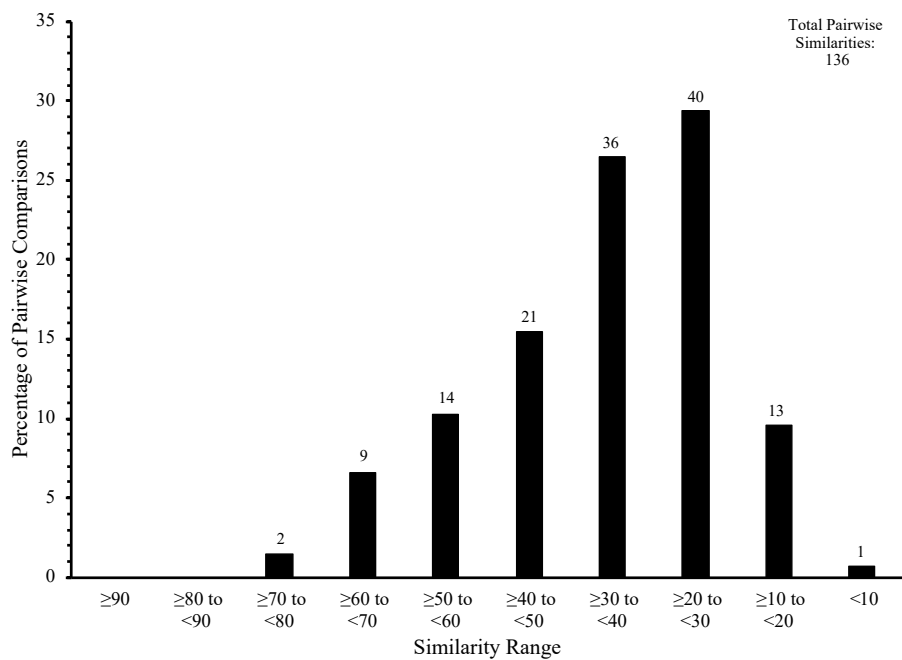


Figure A9. Similarity histogram for moderately absorbed compounds (both training and test set 1 & 2). Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of the pairwise comparisons fall within the 20-50% similarity range.

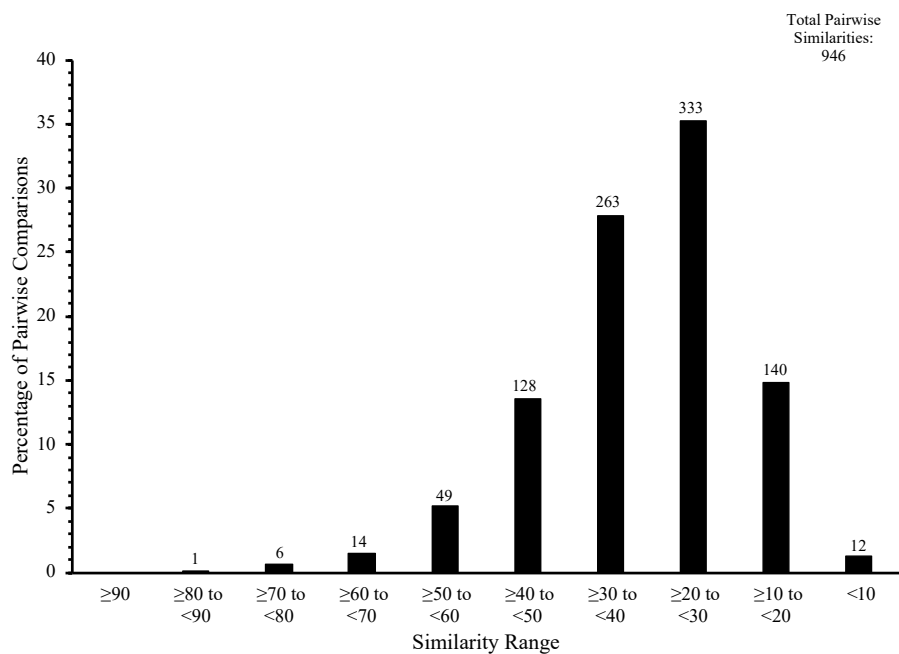


Figure A10. Similarity histogram for well absorbed compounds (both training and test set 1 & 2). Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of the pairwise comparisons fall within the 10-50% similarity range.

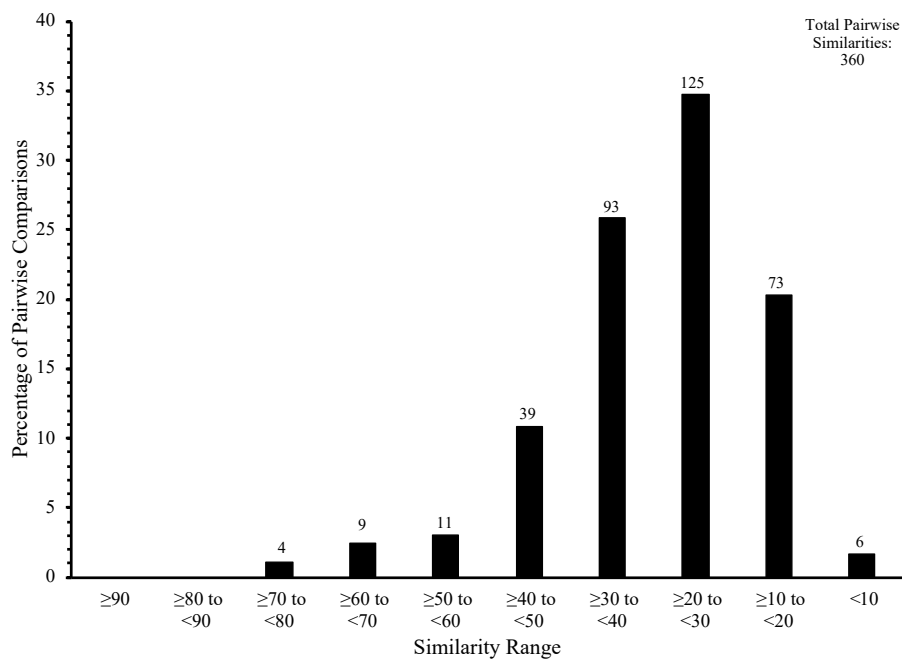


Figure A11. Similarity histogram for test set 2 vs. training set 2 compounds. Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of the pairwise comparisons fall within the 10-50% similarity range.

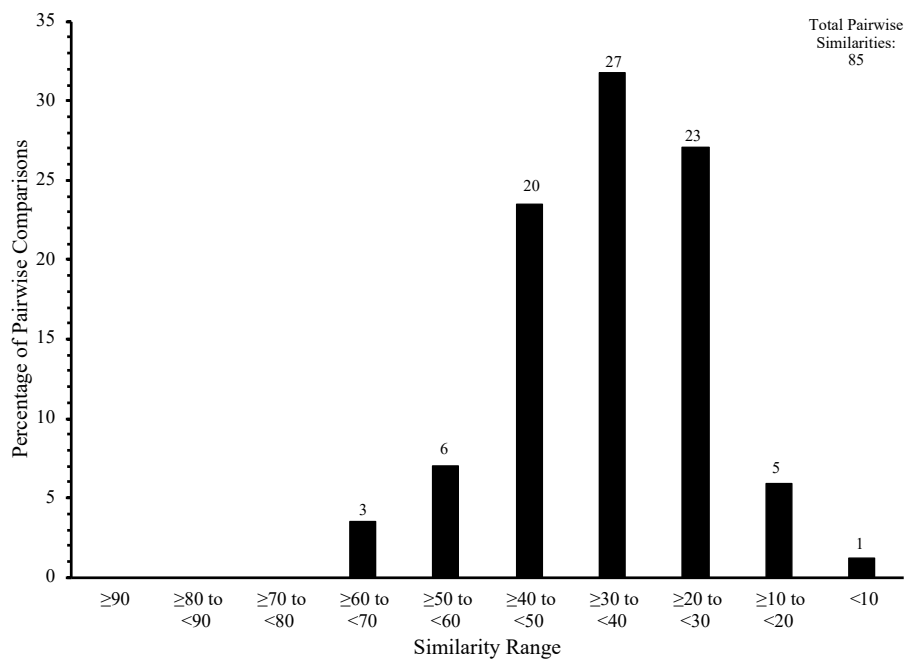


Figure A12. Similarity histogram for poorly vs. moderately absorbed categories (both training and test set 2). Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. All the pairwise comparisons fall within the 0-70% similarity range.

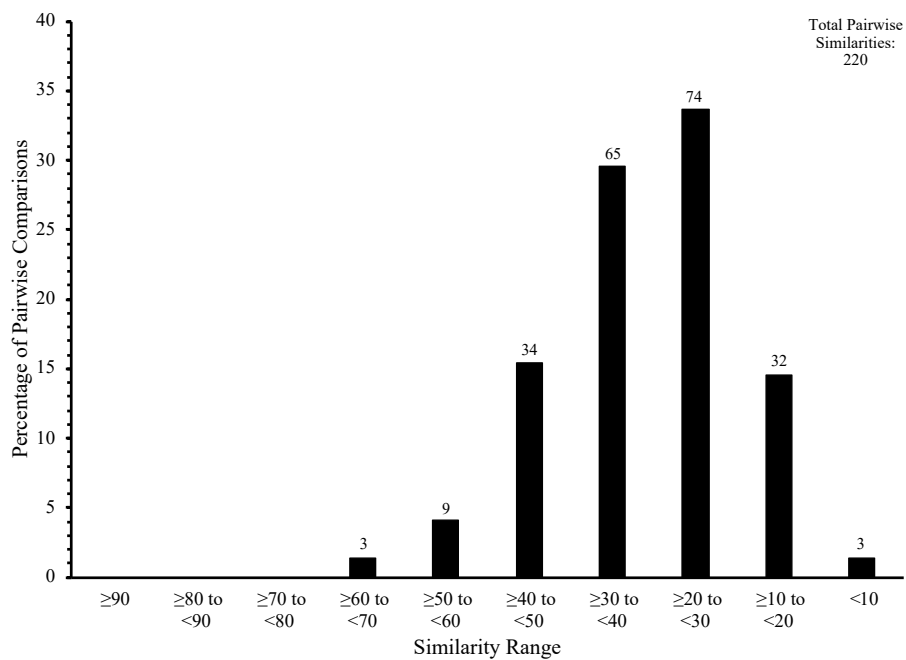


Figure A13. Similarity histogram for poorly vs. well absorbed categories (both training and test set 1 & 2). Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. All the pairwise comparisons fall within the 0-70% similarity range.

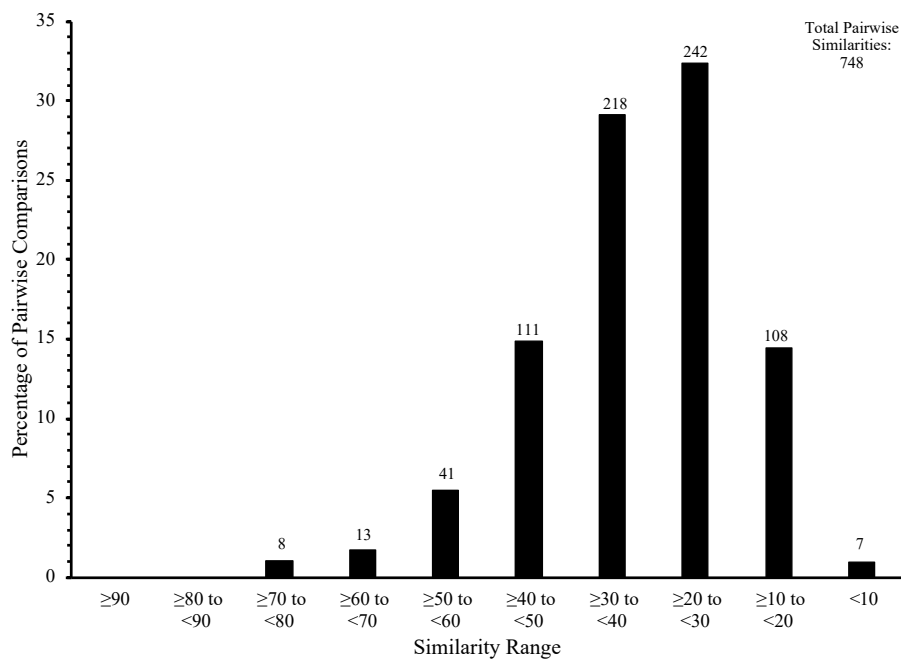


Figure A14. Similarity histogram for moderately vs. well absorbed categories (both training and test set 2). Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of the pairwise comparisons fall within the 10-50% similarity range.

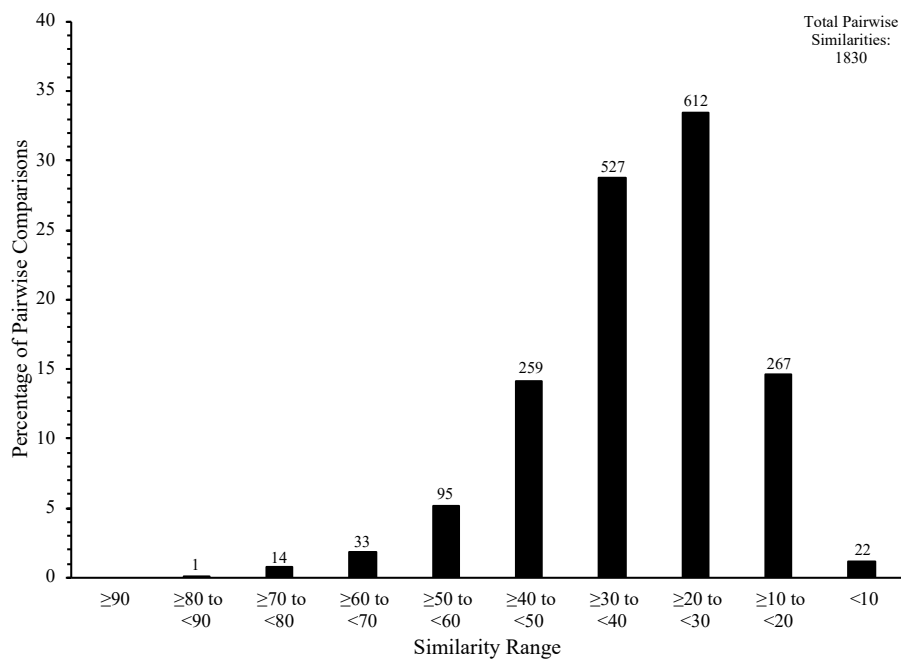


Figure A15. Similarity histogram of all compounds in training and test set 3. Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of pairwise comparisons fall within the 10-50% range.

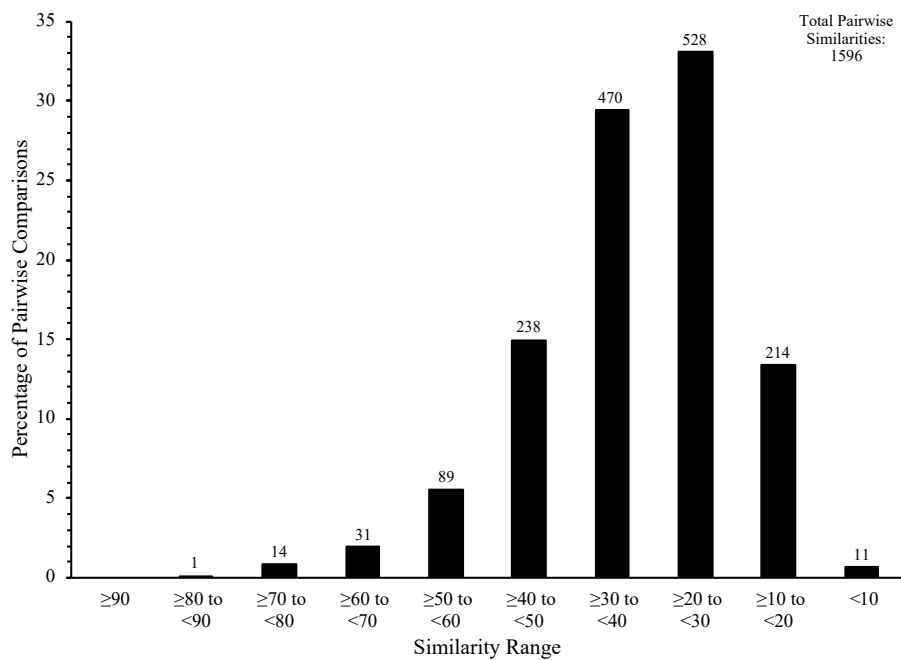


Figure A16. Similarity histogram of training set 3 compounds. Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of pairwise comparisons fall within the 10-50% similarity range.

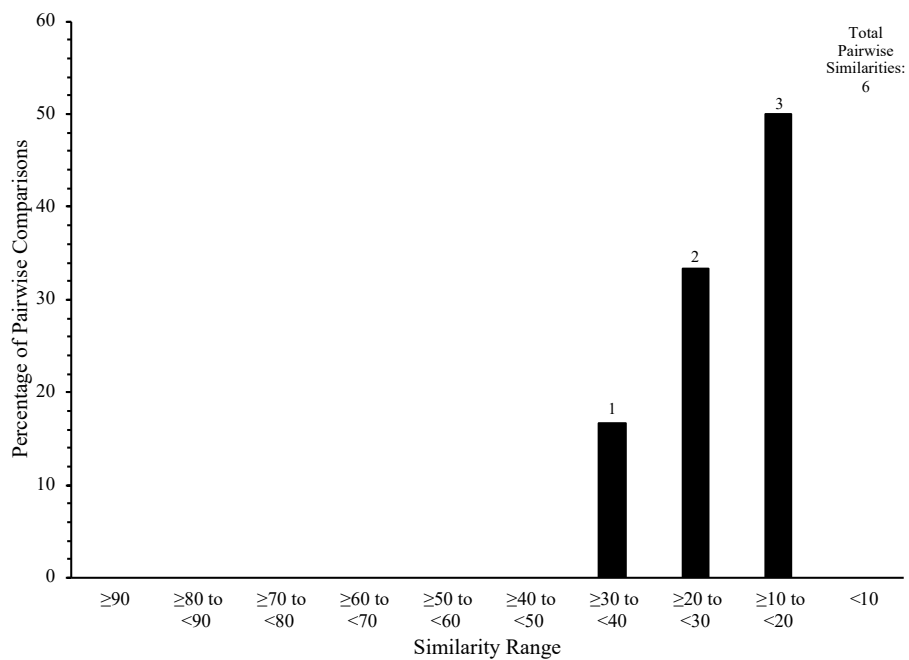


Figure A17. Similarity histogram of test set 3 compounds. Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. All pairwise comparisons fall within the 10-40% range.

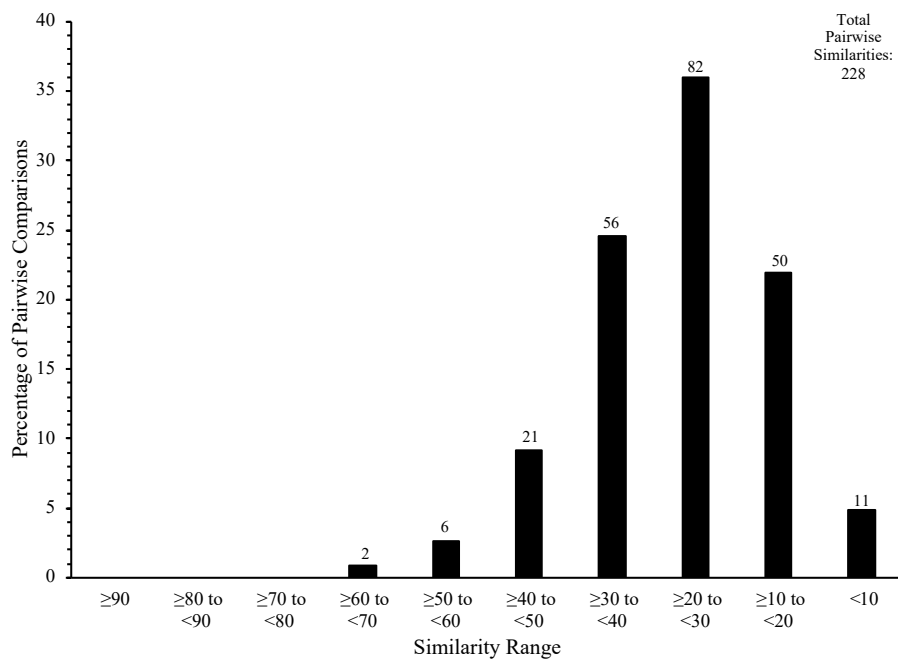


Figure A18. Similarity histogram of test set 3 vs. training set 3 compounds. Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of pairwise comparisons fall within the 10-50% range.

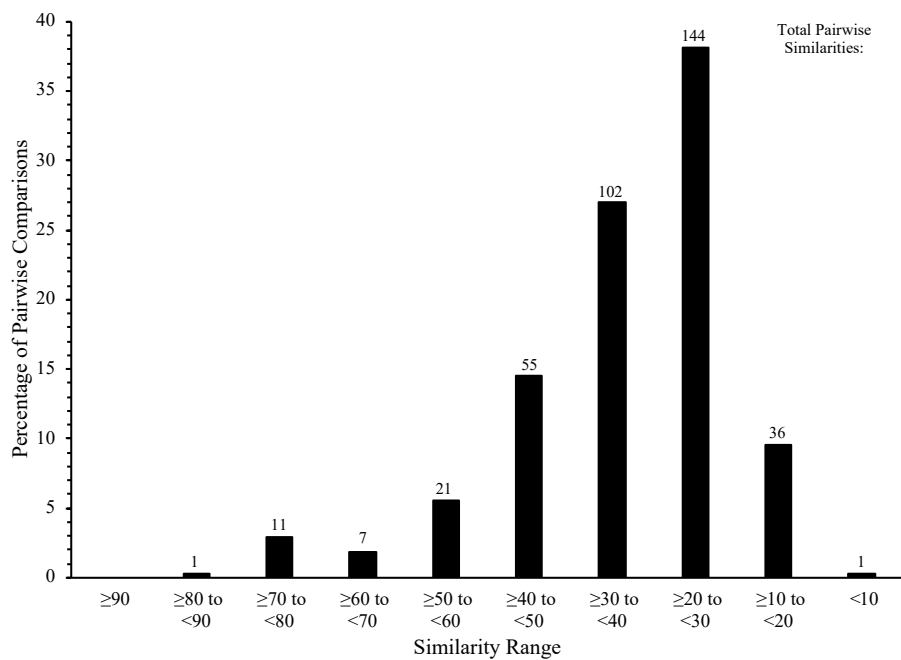


Figure A19. Similarity histogram of low protein binders (both training and test set 3). Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of pairwise comparisons fall within the 10-50% similarity range.

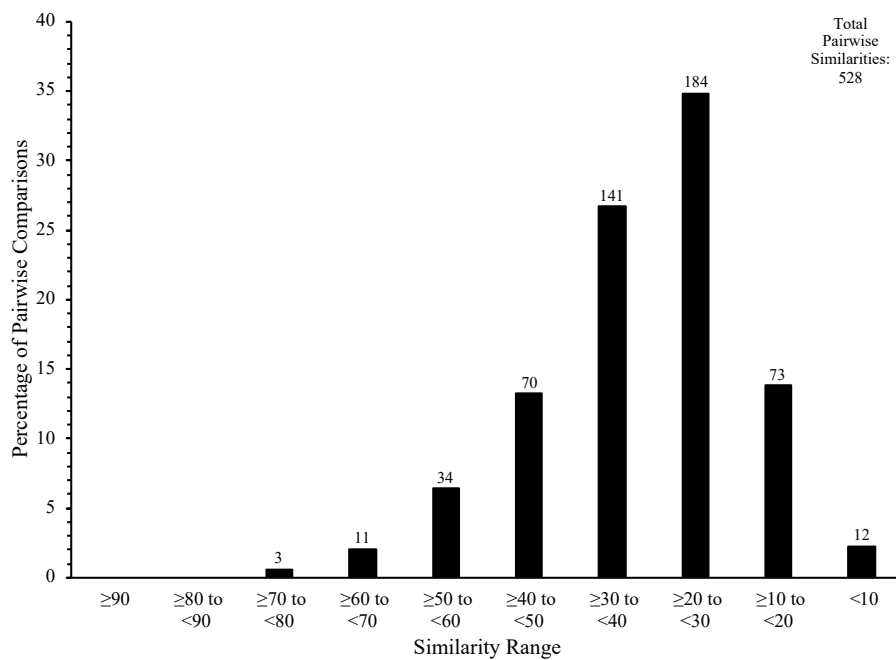


Figure A20. Similarity histogram of high protein binders (both training and test set 3). Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of pairwise comparisons fall within the 10-50% similarity range.

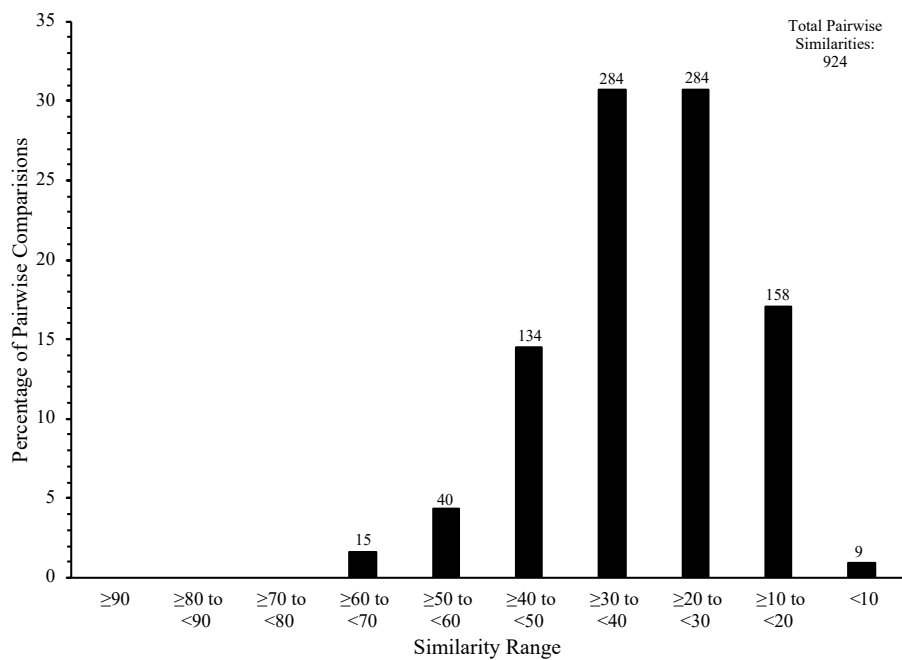


Figure A21. Similarity histogram of low vs. high protein binders (training and test set 3). Generated from similarity matrices of Tanimoto coefficients calculated from MACCS keys of compounds. The majority of pairwise comparisons fall within the 10-50% range.

Table A1. Mispredicted compounds within all 5 HIA models at each stage of model development & validation. Commonly mispredicted compounds indicated in bold.

| | Model 10 (2D) | Model 11 (i3D-gas) | Model 12 (2D3D-gas) | Model 13 (i3D-born) | Model 14 (2D3D-born) |
|--------------------------------|--------------------------|-------------------------------|--------------------------------|--------------------------------|---------------------------------|
| Model Development | mitoxantrone | aztreonam | mitoxantrone | mitoxantrone | mitoxantrone |
| | tranexamic acid | mitoxantrone | tranexamic acid | tranexamic acid | tranexamic acid |
| | ziprasidone | tranexamic acid | ziprasidone | atenolol | ziprasidone |
| | sumatriptan | atenolol | sumatriptan | valsartan | mibefradil |
| | mibefradil | ziprasidone | mibefradil | dipyridamole | fenoterol |
| | fenoterol | mibefradil | fenoterol | ziprasidone | terbutaline |
| | terbutaline | fenoterol | terbutaline | mibefradil | nitrendipine |
| | hydrochlorothiazide | terbutaline | hydrochlorothiazide | erythromycin | acebutolol |
| | nitrendipine | timolol | nitrendipine | terbutaline | timolol |
| | acebutolol | trovafloxacin | acebutolol | ketoconazole | trovafloxacin |
| | timolol | trimethoprim | timolol | trovafloxacin | trimethoprim |
| trimethoprim | | | | | |
| Internal Validation | aztreonam | aztreonam | aztreonam | aztreonam | aztreonam |
| | mitoxantrone | mitoxantrone | mitoxantrone | mitoxantrone | mitoxantrone |
| | tranexamic acid | tranexamic acid | ceftriaxone | ceftriaxone | tranexamic acid |
| | atenolol | atenolol | tranexamic acid | tranexamic acid | atenolol |
| | metformin | ziprasidone | atenolol | atenolol | valsartan |
| | dipyridamole | mibefradil | dipyridamole | valsartan | ziprasidone |
| | ziprasidone | erythromycin | ziprasidone | dipyridamole | mibefradil |
| | furosemide | fenoterol | furosemide | ziprasidone | fenoterol |
| | sumatriptan | terbutaline | sumatriptan | sumatriptan | terbutaline |
| | mibefradil | timolol | mibefradil | mibefradil | nitrendipine |
| | erythromycin | trovafloxacin | erythromycin | erythromycin | acebutolol |
| | fenoterol | trimethoprim | fenoterol | terbutaline | timololo |
| | terbutaline | domperidone | terbutaline | ketoconazole | trovafloxacin |
| | hydrochlorothiazide | | hydrochlorothiazide | nitrendipine | trimethoprim |
| | caffeine | | caffeine | trovafloxacin | domperidone |
| | nitrendipine | | nitrendipine | trimethoprim | |
| | acebutolol | | acebutolol | domperidone | |
| timolol | | timolol | | | |
| meloxicam | | trimethoprim | | | |
| trimethoprim | | domperidone | | | |
| domperidone | | | | | |
| External Validation | lactulose | lactulose | lactulose | lactulose | lactulose |
| | amphotericin b | nadolol | metolazone | nadolol | nadolol |
| | metolazone | | nadolol | | guanabenz |
| | nadolol | | | | |

Table A2. Mispredicted compounds within all 5 Caco-2 permeability models at each stage of model development & validation. Commonly mispredicted compounds indicated in bold.

| | Model 6 (2D) | Model 7 (i3D-gas) | Model 8 (2D3D-gas) | Model 9 (i3D- born) | Model 10 (2D3D- born) |
|--------------------------------|---------------------|----------------------|-----------------------|------------------------|--------------------------|
| Model Development | sulpiride | sulpiride | sulpiride | sulpiride | sulpiride |
| | valsartan | valsartan | valsartan | valsartan | valsartan |
| | sulfasalazine | sulfasalazine | sulfasalazine | sulfasalazine | sulfasalazine |
| | fenoterol | fenoterol | fenoterol | fenoterol | fenoterol |
| | furosemide | furosemide | furosemide | furosemide | furosemide |
| | tranexamic acid | tranexamic acid | tranexamic acid | tranexamic acid | tranexamic acid |
| | erythromycin | erythromycin | erythromycin | erythromycin | erythromycin |
| | terbinafine | atenolol | terbinafine | terbinafine | terbinafine |
| | tamoxifen | terbinafine | tamoxifen | terbutaline | tamoxifen |
| | sulindac | terbutaline | domperidone | acebutolol | sulindac |
| | haloperidol | sumatriptan | haloperidol | tamoxifen | haloperidol |
| | amrinone | acebutolol | amrinone | sulindac | amrinone |
| | morphine | sulindac | morphine | haloperidol | morphine |
| | timolol | haloperidol | timolol | amrinone | timolol |
| | mibefradil | amrinone | ketoconazole | morphine | mibefradil |
| | ketoconazole | morphine | betaxolol | timolol | ketoconazole |
| | betaxolol | timolol | nitrendipine | ziprasidone | betaxolol |
| | nitrendipine | mibefradil | trovafloxacin | mibefradil | nitrendipine |
| | alprenolol | ketoconazole | metoprolol | ketoconazole | alprenolol |
| | trovafloxacin | betaxolol | trimethoprim | betaxolol | trovafloxacin |
| metoprolol | trimethoprim | | naloxone | metoprolol | |
| trimethoprim | omeprazole | | trovafloxacin | trimethoprim | |
| omeprazole | | | trimethoprim | omeprazole | |
| | | | indomethacin | | |
| Internal Validation | ceftriaxone | sulpiride | ceftriaxone | sulpiride | ceftriaxone |
| | sulpiride | valsartan | sulpiride | valsartan | sulpiride |
| | valsartan | sulfasalazine | valsartan | sulfasalazine | valsartan |
| | sulfasalazine | fenoterol | sulfasalazine | fenoterol | sulfasalazine |
| | fenoterol | furosemide | fenoterol | furosemide | fenoterol |
| | furosemide | tranexamic acid | furosemide | tranexamic acid | furosemide |
| | tranexamic acid | erythromycin | tranexamic acid | erythromycin | tranexamic acid |
| | erythromycin | atenolol | erythromycin | hydrochlorothiazide | erythromycin |
| | tamoxifen | terbinafine | terbinafine | terbinafine | hydrochlorothiazide |
| | sulindac | terbutaline | tamoxifen | terbutaline | terbinafine |
| | dipyridamole | sumatriptan | domperidone | acebutolol | tamoxifen |
| | haloperidol | acebutolol | dipyridamole | tamoxifen | sulindac |
| | amrinone | tamoxifen | haloperidol | sulindac | dipyridamole |
| | morphine | sulindac | amrinone | dipyridamole | haloperidol |
| | timolol | haloperidol | morphine | haloperidol | amrinone |
| | mibefradil | amrinone | timolol | amrinone | morphine |
| | ketoconazole | morphine | ketoconazole | mitoxantrone | timolol |
| | betaxolol | timolol | betaxolol | morphine | mibefradil |
| | nitrendipine | ziprasidone | nitrendipine | timolol | ketoconazole |
| | alprenolol | mibefradil | trovafloxacin | ziprasidone | betaxolol |
| trovafloxacin | ketoconazole | metoprolol | mibefradil | nitrendipine | |
| metoprolol | betaxolol | trimethoprim | ketoconazole | alprenolol | |
| trimethoprim | meloxicam | omeprazole | betaxolol | trovafloxacin | |
| omeprazole | trovafloxacin | | naloxone | metoprolol | |
| | trimethoprim | | trovafloxacin | trimethoprim | |

Table A2 Continued. Mispredicted compounds within all 5 Caco-2 permeability models at each stage of model development & validation.

| | Model 6 (2D) | Model 7 (i3D-gas) | Model 8 (2D3D-gas) | Model 9 (i3D- born) | Model 10 (2D3D- born) |
|--------------------------------|-------------------------|------------------------------|-------------------------------|--------------------------------|----------------------------------|
| Internal Validation | | caffeine | | trimethoprim | indomethacin |
| | | indomethacin | | indomethacin | omeprazole |
| | | omeprazole | | ondansetron | |
| | | diclofenac | | | |
| External Validation | metolazone | lactulose | metolazone | lactulose | metolazone |
| | | metolazone | guanabenz | metolazone | |
| | | nadolol | | | |