AUTOMATING PHARMACOKINETIC PREDICTIONS in *ARTEMISIA*

by

Celin Younan

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Biology

Boise State University

May 2023

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Celin Younan

Thesis Title: Automating Pharmacokinetic Predictions in *Artemisia*

Date of Final Oral Examination: 07 March 2023

The following individuals read and discussed the thesis submitted by student Celin Younan, and they evaluated the student's presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Jennifer S. Forbey, Ph.D.	Chair, Supervisory Committee
Sven Buerki, Ph.D.	Member, Supervisory Committee
Lisa Warner, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Jennifer S. Forbey, Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

DEDICATION

I dedicate this work to my amazing family. A special feeling of gratitude to my loving parents, Roubic and Odet Younan whose words of encouragement ring in my ears. My supportive brother Artin and my wonderful Grandma who have never left my side and are very special. Words can hardly describe my thanks and appreciation to you all. You have been my source of inspiration, support, and guidance. You have taught me to be unique, determined, to believe in myself, and to always persevere. I am truly thankful and honored to have you in my life.

ACKNOWLEDGMENTS

Many individuals made this thesis research possible. First and foremost, I thank my graduate committee for their support, encouragement, and diverse prospective on research. Words cannot express my gratitude to my professor and major advisor Dr. Jennifer Forbey for her invaluable patience, feedback, and introducing me to the field of chemical ecology, while also acquainting me with a community of people who were able to help me connect chemical ecology with my passion for computational biology. I also could not have undertaken this journey without my defense committee, Dr. Sven Buerki and Dr. Lisa Warner, who generously provided knowledge and expertise to help me along this journey.

I also want to thank other faculty with whom I have interacted at Boise State University. In particular, Eric Hayden, Amy Ulappa, and Carolina Viera who added to the quality of my graduate student experience. My interactions with these professors have come at different times and at different capacities, but each have had an influence on my graduate career. I also want to thank the many graduate and undergraduate students who I have experiences with both in and out of the classroom who have made my graduate experience more enjoyable.

A special thanks to Dr. Stephanie Galla, who has been an amazing mentor as well as providing images that were used in this thesis, Britt Pendleton for also providing amazing images that were used, and Jim Beck, who guided me through the end of my

undergraduate studies and into my graduate career and has helped me throughout my thesis project.

Many members of the Forbey lab, both past and current, have helped me through my thesis research. Debbie Connor, Britt Pendleton, and Bryanna Bright provided assistance and support throughout my graduate program.

Thank you to Boise State University and the Department of Biological Sciences specifically the support from the graduate program, as well as TA support.

This thesis was made possible by NSF Track 2 EPSCoR Program under award number OIA-1826801 and the NSF Track 2 Idaho EPSCoR program under award number OIA-1757324.

ABSTRACT

Pharmacokinetics (PK) is the time course of a compound in the body that is dependent on mechanisms of absorption, distribution, metabolism, and excretion or ADME. A thorough understanding of PK is essential to predict the consequences of organisms exposed to chemicals. In medicine, predictions of PK of drugs allows us to properly prescribe drug treatments. In toxicology, PK allows us to predict the potential exposure of environmental contaminants and how they may affect organisms at the time of exposure or in the future. Chemical ecology could benefit from computational predictions of PK to better understand which plants are consumed or avoided by wild herbivores. A limitation in computational predictions of PK in chemical ecology is the large quantities of biodiverse natural products involved in complex plant-herbivore-microbial interactions compared to biomedical and environmental toxicology studies that focus on a select number of chemicals. The objective of this research was to automate the process of mining predicted PK of known chemical structures in plants consumed by herbivores and to use predicted PK output to test hypotheses. The first hypothesis is that because monoterpenes are smaller in molecular weight and have relatively high lipophilicity when compared to phenolics and sesquiterpenes, they would have higher absorption, be more likely to be substrates for efflux transporters that regulate absorption, and be more likely to inhibit metabolizing enzymes than phenolics and sesquiterpenes. The second hypothesis is that monoterpenes that are induced or avoided by foraging herbivores would have higher absorption, be less likely to be substates for efflux

transporters, and be more likely to inhibit metabolizing enzymes compared to the individual monoterpenes that are not induced or avoided by herbivores. This automated approach used Python packages to obtain chemical notations from the PubChem website and mine predicted PK information for chemical input from the SwissADME website. The PK output from SwissADME was analyzed using ANOVAs to test for differences in molecular weight and lipophilicity among chemical classes (monoterpenes, phenolics, and sesquiterpenes). Chi-squared tests were used to assess if chemical groups had high or low absorption, were substrates of efflux transporters, or inhibited metabolizing enzymes. Mined PK data for chemicals can be used to understand drug-drug interactions in pharmacology, predict exposure to environmental contaminants in toxicology, and identify mechanisms mediating plant-microbe-herbivore interactions. However, the broad benefits of mining predicted PK across disciplines requires a workforce with competency in chemistry, physiology, and computing who can validate the automation process and test hypotheses relative to different disciplines. Course-based and Lab-based Undergraduate Research Experiences (CUREs and LUREs) have been proven to not only improve grades but also increase engagement diversity and inclusion. As a graduate teaching assistant, I created and taught a PK LURE module in an undergraduate Animal Physiology and Nutrition course to create a sustainable quality control step to validate input of chemical structures and PK output generated from the automated process. The course simultaneously provided students with an authentic research experience where they integrated chemistry, pharmacology, computing, public databases, and literature searches to propose and test new hypotheses. Students gained indispensable interdisciplinary research skills that can be transferred to jobs in veterinary and human

medicine, pharmaceutics, and natural sciences. Moreover, undergraduates used existing and new PK data to generate and test novel hypotheses that go beyond the work of any single graduate student or discipline. Overall, the integration of computing and authentic research experiences has advanced the research capacity of a diverse workforce who can predict exposure and consequences of chemicals in organisms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PICTURES

LIST OF ABBREVIATIONS

PK	Pharmacokinetics
ADME	Absorption, Distribution, Metabolism, Excretion
PSM	Plant Secondary Metabolite
PCB	Polychlorinated Biphenyls
GI	Gastrointestinal
CYP	Cytochrome P450
P-gp	P-glycoprotein
SMILES	Simplified Molecular Input Line Entry System
HTML	Hypertext Markup Language
CSS	Cascading Style Sheets
CID	Compound ID
CSV	Comma-Separated Values
PDF	Portable Document Format
IP	Internet Protocol
VPN	Virtual Private Network
API	Application Programming Interface
CURE	Course-based Undergraduate Research Experience
LURE	Lab-based Undergraduate Research Experience
VIP	Virtually Integrated Project
GPA	Grade Point Average

REU	Research Experience for Undergraduates
KSA	Knowledge, Skills, and Abilities
STEM	Science, Technology, Engineering, Mathematics

CHAPTER ONE: AUTOMATING PHARMACOKINETIC PREDICTIONS: USING *ARTEMISIA* AS A CASE STUDY

Abstract

Pharmacokinetics (PK) is the study of what the body does to a metabolite and is important to chemists, biologists, and pharmacologists. Obtaining predicted PK from chemical structures is often embedded deep within websites and requires a multi-step retrieval process to access data manually. However, data retrieval from websites creates a bottleneck in mining predicted PK when a large number of metabolites are being considered as is often the case for biodiverse natural products. The automation of this multi-step process would save time and allow researchers to support reproducible science in this field and also allow for more accurate results while retrieving PK data from diverse sources of metabolites that can be used to test hypotheses. The objective of this research was to automate the process of mining PK data using known chemical structures in plants consumed by herbivores and to use PK output to test chemical ecology hypotheses as a case study. We used a list of chemicals from the North American endemic *Artemisia* species within the subgenus *tridentatae* (hereafter, sagebrush) as a test case to automate a reproducible pipeline of mining PK data from publicly available websites. Sagebrush offers an ideal case study because it: 1) contains a large number of metabolites within several distinct classes of chemicals that have diverse chemical properties that vary in composition and concentration within and among species; 2) has parent compounds that are known to be metabolized into new chemicals by animal and

microbial enzymes which further increases structural diversity of chemicals; and 3) has chemicals with known pharmaceutical properties and known interactions with herbivores. We used the automated retrieval process to mine PK values of 166 metabolites in sagebrush and used PK output to test biological hypotheses. First, we investigated patterns in physicochemical properties of the three main classes (monoterpenes, phenolics, sesquiterpenes) of chemicals in sagebrush. Using a one-way ANOVA, we found that the molecular weight of monoterpenes was significantly lower than phenolics and sesquiterpenes and that monoterpenes and sesquiterpenes were more lipophilic than phenolics. We used chi-squared tests to assess the hypothesis that smaller size and relatively high lipophilicity of monoterpenes would increase gastrointestinal (GI) absorption and therefore increase interactions with the efflux transporter P-glycoprotein (P-gp) and the metabolizing enzymes cytochrome P450 (CYP). We found no difference in GI absorption among the three chemical classes. However, sesquiterpenes were most likely to be P-gp substrates and monoterpenes and sesquiterpenes were less likely than phenolics to be CYP inhibitors. This suggests that sesquiterpenes may be the least bioavailable because they are effluxed out of cells by P-gp but could also be P-gp inhibitors or have interactions with other transport proteins. Results also suggest that phenolics may influence PK of co-occurring chemicals through CYP inhibition, which could cause toxicity. We also assessed if our new PK output could explain observed plant-herbivore interactions. Specifically, we tested the hypothesis that monoterpenes in sagebrush avoided by wild herbivores and those induced by damage from herbivores would have higher absorption and be more likely to interact with CYPs compared to the individual monoterpenes that are not induced or avoided by herbivores. Using chi-

squared and Tukey's HSD tests we found that monoterpenes that are induced or avoided by foraging herbivores had lower GI absorption, but did not differ in interactions with P-gp or inhibition of CYPs compared to monoterpenes that are not induced or avoided by herbivores. This suggests that the site of action for monoterpenes that are induced and avoided may be in the intestine of herbivores. We demonstrate the value of automating the multi-step process of mining predicted PK data to test biological hypotheses (See Appendix A). Results of the PK hypotheses can generate predictions that can be empirically tested using *in vitro* and *in vivo* experiments to confirm absorption and interactions with P-gp and CYPs and predict the overall exposure of ingested chemicals in herbivores or to target the discovery of bioavailable natural products for medicine. Moreover, our process for mining PK can be expanded to include structures of unidentified chemicals generated using analytical chemistry that arise from animal and microbial metabolism of parent chemicals from plants.

Introduction

Predicting changes in concentrations of bioactive chemicals in an organism and how those chemicals influence the physiological function of organisms is essential to understanding and managing the health of humans (Kay, 2006), domestic species (Rajaganapathy et al., 2011), and non-model organisms (Maurya et al., 2019). Pharmacokinetic (PK) modeling has been beneficial since its inception to predict exposure (Leven et al., 2019) and resistance to drugs (Pathania et al., 2018) and optimize dosing regimens (Lucas et al., 2019) using multiple parameters. Pharmacokinetics can be described as how the body affects a compound, molecule, or metabolite, which refers to how the body absorbs, distributes, metabolizes, and excretes the metabolite, also known

as ADME (Eddershaw et al., 2000). As such, PK is used to predict the time course of an ingested metabolite in the body (Wishart, 2007).

Predicting ADME from chemical structures offers pharmacological advantages in drug discovery and development that benefits human health and can be used to predict toxic consequences of environmental contaminants. In pharmacology, PK predictions are used to improve clinical outcomes. For example, PK can predict the correct dosing of morphine relative to the age of young children to achieve pain relief (Verscheijden et al., 2021) and can explain variable pain relief associated with renal failure in patients (Mazoit et al., 2007). PK data can also be used to predict if parent compounds from natural products are more bioavailable than their synthetic counter-parts (e.g., Paclitaxel compared to Docetaxel, Sharifi-Rad et al., 2019) or if detoxification products are more bioactive than the parent compound (e.g., curcumin compared to hexahydrocurcumin, Huang et al., 2018). PK can also be used to predict drug-drug interactions where one drug influences the kinetics and therefore exposure of another co-administered drug through inhibition of the proteins that regulate absorption and metabolism of chemicals (Percha & Altman, 2013). For example, drug-drug interactions can cause toxicity of sildenafil if not properly co-administered with *Bancha* tea methylxanthines (Radeva-llieva et al. 2022). This is because methylxanthines in green tea inhibit the cytochrome P450 enzyme (CYP1A2) and this inhibition reduces the rate by which CYP1A2 metabolizes sildenafil resulting in an excess amount of sildenafil in the blood which causes toxicity (Radeva-llieva et al. 2022). Finally, PK models (Ahmad, 2007) are used by toxicologists to estimate the potential exposure of environmental contaminants and how they may affect organisms at the time of exposure or in the future. For example, several studies use PK to

predict exposure to Polychlorinated Biphenyls (PCBs, Parham et al., 1997, Lutz et al., 1984) which are highly toxic industrial compounds that can accumulate in the environment (Bagale, 2022) and have dose-dependent health consequences for animals (Krause et al., 2022) and humans (Zhu et al., 2022). Overall, PK models have been critical in advancing drug development and predicting therapeutic and toxic consequences of exposure to natural products, their derivatives, and synthetic molecules.

Unlike numerous examples found in human medicine and environmental toxicology, there are relatively fewer examples of using PK in chemical ecology to predict how wild animals interact with natural products in their diet (although see Mazorra-Alonso et al., 2021, Godfray et al., 2019, Dyer et al., 2018, Freeland & Janzen, 1974). Plants often produce metabolites to defend themselves against herbivores that consume them (Mazid et al., 2011). The natural products synthesized by plants exist in complex mixtures and together they are known as plant secondary metabolites (PSMs). Chemical ecologists, like pharmacologists and toxicologists, can use PK to better understand the larger numbers of metabolites arising from plant-herbivore-microbial interactions. Specifically, the plants herbivores consume contain nutrients as well as potentially toxic (Takahashi & Shimada, 2008) or therapeutic (Díaz-Navarro et al., 2021) PSMs. Predicting the fate of ingested PSMs in the body will help ecologists understand selection or avoidance of PSMs by herbivores (Ulappa et al., 2014). There is some evidence that the pharmacokinetics of PSMs, including absorption (Sorensen et al., 2004, Sorensen et al., 2005, K. Kohl & Dearing, 2017), rates of metabolism (McLean et al., 2007, Nobler et al., 2019), and the types of metabolites excreted (Staudenmaier et al., 2022, K. D. Kohl et al., 2018) do explain the foraging behavior (i.e., dosing regimen) of

vertebrate herbivores. For example, higher absorption of α-pinene by generalist herbivores results in lower excretion of α-pinene unchanged in the feces compared to specialist herbivores (Sorensen et al., 2004) and may explain higher blood concentrations of α-pinene after an oral dose (Sorensen et al., 2003) and lower intake of food with α-pinene (Sorensen et al., 2005) by generalists compared to specialists. Similarly, the metabolites of p-cymene differ in oxidation in specialist and generalist herbivore species (Boyle et al., 1999) and may explain differential tolerance to plants that vary in p-cymene among herbivores (Massing et al., 2021). However, given the logistical and ethical challenges of *in vitro* and *in vivo* PK studies in wildlife, there is a need for new approaches to predict the fate of ingested PSMs and advance our understanding of the chemical ecology of plant-herbivore interactions.

One approach to better understand the chemical ecology of plant-herbivore interactions, is to predict PK using the chemical structure of PSMs in forage consumed or avoided by herbivores (Forbey et al., 2013). Specifically, a chemical's three-dimensional structure can be translated into a string of chemical symbols that are interpreted using computer software programs to generate a single canonical form, called the Canonical SMILES (Simplified Molecular Input Line Entry System) notation. This notation encodes for 3D chemical information into a string system and stores complex information into a simple system that can easily be used as a precise input for research purposes. The chemical structures generated by analytical analysis (e.g., mass spectrometry or Nuclear Magnetic Resonance [NMR], Liu et al., 2019, Seger & Sturm, 2022) of extracts from plants or excreta from non-model organisms that cannot be identified using databases (i.e., PubChem, Kim et al., 2021) can still be used to generate Canonical SMILES. Once

the Canonical SMILES is generated for each chemical, a multi-step retrieval process can be used to mine predicted ADME of that chemical. This retrieval process is normally done manually, because of the lack of an Application Program Interface (API). First, the Canonical SMILES notation of a target chemical is obtained by entering the name of the target chemical into a search bar on the PubChem website, then navigating to the chemical, and finally obtaining the Canonical SMILES notation. Once the researcher has the Canonical SMILES, they must navigate to the SwissADME website (Daina et al., 2017) which is a database that uses Canonical SMILES notation as an input to mine PK data for the associated chemical. Physiochemical data and ADME predictions are generated for each Canonical SMILES entered into a search bar.

The manual retrieval of parameters from multiple databases (PubChem and SwissADME) creates a bottleneck in mining PK data when considering studies investigating large numbers of metabolites found in natural products. Pharmaceutical scientists typically focus on predicting PK using *in silico* and experimental methods (Lombardo et al., 2017) from a limited number of chemicals associated with a single drug and detoxification products. In contrast, chemical ecologists often work with mixtures of PSMs that may include hundreds of individual chemicals representing diverse chemical classes (Lautié et al., 2020, Karunanithi & Zerbe, 2019), many of which are not found in databases (Mushtaq et al., 2018, Khalifa et al., 2019). Complexity is further increased from plant-herbivore (Li et al., 2022) and plant-microbial (Carvalhais et al., 2013) interactions which involve multiple and often simultaneous enzymatic biosynthesis of parent chemicals and detoxification of parent chemicals into new metabolites. When dealing with hundreds of parent chemicals and metabolites, the manual retrieval process

for obtaining ADME information to predict PK not only becomes incredibly time consuming, but also prone to error.

To address this need, we used Python (Van Rossum and Drake, 2009) and associated packages to automate the multi-step process to accurately and effectively retrieve Canonical SMILES and predictive PK data from large data sets. In this case study, we demonstrate the process of determining the PK of known chemical structures in plants consumed by herbivores by investigated PSMs found in the North American endemic *Artemisia* species. The main PSMs in sagebrush are terpenoids, flavonoids, coumarins, caffeoylquinic acids, sterols, and acetylenes (Bora & Sharma, 2011). Sagebrush offers an ideal case study because it contains diverse classes of PSMs, with a large number of metabolites within each class that co-occur and vary across species and subspecies (Turi et al., 2014). PSMs in sagebrush vary in presence, concentration, and bioactivity. Sagebrush is also a potential source of parent mixtures of PSMs that may result in amplified biodiversity of metabolites generated from metabolizing enzymes in herbivores and microbes that interact with this plant. Sagebrush is part of a taxonomic group with historical and current pharmacological properties and therefore a source of potential biomedical insights (Ekrami et al., 2022; Kelley et al., 1992; Sánchez et al., 2010). While *Artemisia* species are used as medicines all over the world, ADME adaptations of herbivores to tolerate these mixtures are poorly understood (Nobler et al., 2019, Oh et al., 2019). We used the sagebrush system as an example of obtaining PK data for a diverse set of chemicals, using coding techniques to mine PK data directly from a database that does not offer an API, and using Python as a potential solution for interpreting data.

We used the databases PubChem (Kim et al., 2021) and SwissADME (Daina et al., 2017) combined with Python and associated packages to automate the process of conducting searches on multiple databases to access both chemical structure and mine ADME parameters. The main objective of this research was to create an automated pipeline that first assembles a list of PSMs (*.csv), extracts the structures of those PSMs from PubChem, then inputs the structures of PSMs into SwissADME to mine and organize ADME parameters for each metabolite. The ADME parameters can then be used to predict PK in organisms interacting with PSMs. This pipeline can be used as a foundation to create a similar pipeline that interacts with the SwissTargetPrediction database (Gfeller et al., 2014) to mine for predicted molecular targets of PSMs and their metabolites.

The secondary objective of this research was to use the mined ADME output to identify patterns in the physiochemical properties of the three main classes of PSMs (monoterpenes, phenolics, and sesquiterpenes, Turi et al., 2014) in sagebrush species and test chemical ecology hypotheses. We tested the hypothesis that because monoterpenes are smaller in molecular weight (g/mol) and have high lipophilicity, they would have higher gastrointestinal (GI) absorption and would be more likely to be P-glycoprotein (P-gp) substrates and cytochrome P450 (CYP1A2) inhibitors than phenolics and sesquiterpenes. Second, we hypothesized that within monoterpenes, those that are induced or avoided by foraging herbivores would have higher GI absorption, would be less likely to be P-gp substrates, and would be more likely to be CYP1A2 inhibitors compared to individual monoterpenes that are not induced or avoided by herbivores.

Implementation and Methods

The automated pipeline for mining PK output of PSMs consists of five main steps (Figure 1.1) after completing step 0) Getting started with important preliminary steps. The main steps include: 1) Generate chemical data from diverse sources; 2) Import and prepare chemical data; 3) Retrieve Canonical Simplified Molecular-Input Line-Entry System (Canonical SMILES) from PubChem; 4) Retrieve ADME information from SwissADME; and 5) Application of the automated pipeline using sagebrush chemistry as a case study to test chemical ecology hypotheses using statistical tests. Steps 1-4 are combined in our pipeline (GitLab repository, see Appendix A) to produce information about each chemical that will generate PK output for each individual PSM in animals that consume chemically defended plants.

Operating System, Programs, and Packages

This pipeline has been tested on iMac running macOS version 11.4 using Python version 3.7.13. Packages used include Pandas (McKinney, 2010), Numpy (Harris et al., 2020), pubchempy (https://pypi.org/project/PubChemPy/1.0/), geckodriver, selenium, csv (Shafranovich, 2005), and requests (Chandra & Varanasi, 2015). We used Firefox web browser version 108.0.2. All statistical analyses were conducted using JMP Pro 16.1 (SAS Institute Inc. 2021). All plots were made in R version 4.1.0 using ggplot2 (Wickham, 2016).

Step 0. Getting Started with Important Preliminary Steps

Prior to initiating this pipeline, we must have a list of chemicals ready to analyze, have a file structure in place, and have very basic web development knowledge.

Original Chemical List

The original chemical list provided the names of 204 individual PSMs identified in the North American endemic A*rtemisia* species within the subgenus *tridentatae* obtained from Turi et al. (2014). The chemical names were placed manually into a spreadsheet which was loaded into Python as a CSV file. The datasheet was formatted to only include the column named 'Chemical name' and this column was extracted and placed in a list called 'ChemList_Column' using an underscore to removing spaces. These minor changes improved the quality of our data output for subsequent steps.

File Structure

Within the repository there are three folders (Appendix A): Data, Code, and Output. Within the Data folder there is one CSV file that contains the raw data used for this project. The Code folder contains one file that contains the pipeline written in Python. The Output folder contains the output produced from the automated PK pipeline after it is run.

Web Development

HyperText Markup Language (HTML), Cascading Style Sheets (CSS), and JavaScript were the web development tools used to efficiently extract ADME data from the SwissADME database. HTML dictates the structure of any website. CSS refers to the style of a website and dictates the text size, color, font, etc. The JavaScript of the website controls the web function. Understanding how the JavaScript of the SwissADME database is structured was critical for writing a function that could manipulate the database to output ADME data.

Step 1. Generated Chemical Data From Diverse Sources

In the field of chemical ecology, data can be produced from diverse sources. Possible sources include: plant tissue for parent chemicals (e.g., (Sánchez et al., 2010)); microbial cultures inoculated with parent chemicals that produce new metabolites (e.g., (Kohl et al., 2018)); and animal metabolizing enzymes inoculated with parent chemicals that produce new metabolites (e.g., (Forbey et al., 2018)). For our purposes we used plant chemical names produced in tissue that we obtained from a single review paper (Turi et al., 2014).

Step 2. Import and Prepare Chemical Data

We used chemical names from tables found in Turi et al. 2014 (Table 3, 4, and 5) as a case study to test our pipeline. We specifically used the chemical names because using the chemical formula may result in multiple chemicals that have the same chemical formula (conformational isomers).

The automated PK pipeline started with importing a data file (spreadsheet) in CSV format. In our case study, the CSV data file contained a list of names of PSMs found in sagebrush that were identified from existing literature (e.g., Turi et al., 2014). The names of each PSM (Chemical names column) were converted into a list format and used as an input to retrieve the Canonical SMILES data for each chemical.

Step 3. Retrieve Canonical Simplified Molecular-Input Line-Entry System (Canonical SMILES) from PubChem

We used PubChem (Kim et al., 2021) to obtain the Canonical SMILES data for each chemical in our list which will later be used as input in the SwissADME website to mine PK output. We accessed PubChem using a package developed for Python called

pubchempy, which works through Compound Identifiers (CIDs) to obtain Canonical SMILES for chemicals it can identify. This step was completed using two main loops. The first loop obtained the CIDs by iterating over the chemical names, conducting a PubChem search through pubchempy, and placing the CIDs that were found into a list ('foundcid') and placing the chemical names that were not found in the search into a separate list ('unfoundcid'). The second loop used the CIDs from the 'foundcid' list to conduct a PubChem search using pubchempy to obtain the molecular formula, molecular weight (g/mol), and Canonical SMILES for each specific chemical. For this loop, we also created empty lists for each of the searches conducted. The CIDs, molecular formula, and molecular weight are important parameters included in our final output data sheet, but only the Canonical SMILES are needed to obtain ADME information.

Step 4. Retrieve Absorption, Distribution, Metabolism, and Excretion (ADME) information from SwissADME

The automated next step in our PK pipeline involves retrieving the ADME parameters from the SwissADME database. SwissADME has multiple ways of generating PK data for the user including in-house Support Vector Machine (SVM) models, data obtained from publications, or data directly computed through non-commercial executables. All of which introduce varied levels of uncertainty. We used the selenium package, an automation tool, along with multiple feature capabilities the package contains. A Firefox browser was opened and navigated to the SwissADME database. First, the pipeline navigated to the input location of the Canonical SMILES and used a loop to input a Canonical SMILES. Once the Canonical SMILES were inputted into the SwissADME database, it navigated to and clicked the Run button. Once the run

is complete for a Canonical SMILES, the ADME data is imported as a CSV file link as a hypertext reference (href) attribute into a list of CSV file links. Finally, the browser is closed.

The SwissADME database has multiple options for viewing output data for each targeted chemical that is used as input. For our case study, we chose to use comma-separated values (CSV) file format to retrieve data. Using the browsers web inspector functionality, we were able to inspect the JavaScript code of the database and determined that the output can be obtained as a href link. The href link for the chemical list (with Canonical SMILES) and was finally read using the pandas read_csv function on Python.

The CSV file link was downloaded. Once the data was downloaded, it was converted into a pandas data frame for a more workable set of data. This data frame included the chemical name, molecular formula, molecular weight (g/mol), Canonical SMILES, and a subset of ADME parameters. There are 46 parameters generated by the SwissADME database, but we focused on a subset of these parameters required to test specific chemical ecology hypotheses. Merging this information created a detailed output ADME file with all the information for our known list of chemicals (GitLab repository in the “Data” folder, see Appendix A).

We validated the automated input and output of our PK pipeline using undergraduate Biology students who manually generated ADME parameters for chemicals and compared manual output to automated output as part of a lab module in a Lab-Based Undergraduate Research Experience (LURE) focused on Animal Physiology and Nutrition (see Chapter 2). Students helped test for accuracy of the pipeline using the same chemical lists we tested (monoterpenes, phenolics, and sesquiterpenes, Turi et al.,

2014). Students were able to identify a subset of issues that led to solutions to problems we were having with our output (Table 1.1). This validation process helped us identify errors and revise the pipeline to deal with these errors (Table 1.1).

Step 5. Application of the automated pipeline using sagebrush chemistry as a case study to test chemical ecology hypotheses using statistical tests

Once all of the metadata for the list of PSMs was obtained and validated, the CSV file was imported into JMP to conduct statistical analysis to test our chemical ecology hypothesis. We focused on several ADME parameters that were most likely to influence PSM exposure following intake by an herbivore. The physiochemical parameters of interest included molecular weight (g/mol) and relative lipophilicity and the ADME parameters of interest included GI absorption and whether the chemical was a substrate for P-gp or a CYP1A2 inhibitor. We chose molecular weight because chemicals with a molecular weight less than 400 g/mol are more likely to have high bioavailability (Ma et al., 2021). We chose lipophilicity because higher lipophilicity increases absorption (Alavijeh et al., 2005) and decreases solubility and metabolic stability (Parrott et al., 2022). We chose GI absorption because although the intestine is not the only route of absorption, PSMs consumed by herbivores must be absorbed by enterocytes to reach the systemic circulation (Williams et al., 2022). We chose whether the chemical was a P-pg substrate because binding affinity to P-gp plays a role in absorption, distribution, and excretion (Ma et al., 2021). We chose whether the chemical was a CYP1A2 inhibitor because inhibition and induction of this metabolizing enzyme is associated with drug-drug interactions and can cause serious adverse physiological consequences (Kato, 2020). A one-way ANOVA was used to analyze the physiochemical properties (dependent

variables: molecular weight [g/mol] and lipophilicity) of the three classes (independent variables: monoterpenes, phenolics, and sesquiterpenes) of chemicals in sagebrush. To evaluate differences between classes of chemicals, we followed significant results with pairwise comparisons using a Tukey's HSD test adjusted p-value. We used non-parametric chi-squared tests to test categorical differences in the dependent variables (GI absorption [high versus low], P-gp substrate [yes versus no], and CYP inhibition [yes versus no]) between the independent variable of chemical classes (monoterpenes, phenolics, and sesquiterpenes) and between categories within the monoterpenes (e.g., induced/avoided and not induced/avoided). For each of these tests we set alpha as 0.05.

Results

PubChem: Generating Canonical SMILES

Quality control of our output of obtained CIDs by students (see Chapter 2) revealed that 24.9% (53/204) of chemicals from our original list did not have CIDs on PubChem. Inability to obtain CIDs for these 53 chemicals included having multiple chemical names in a single cell on a spreadsheet, misspelling of the chemical name, or because PubChem used a number instead of the chemical name (e.g., Chemical Name on PubChem was 10180-88-8 instead of our Chemical Name of Deacetoxymatricarin). We were able to manually correct these issues for 14 of the 53 chemicals that were not found on PubChem resulting in CIDs for 166 of 204 individual chemicals that have been identified in sagebrush species in North America (81.37% of original list).

Retrieval of Absorption, Distribution, Metabolism, Excretion (ADME) data

While automating the ADME retrieval step, we ran into an issue with the SwissADME database stating there were 'too many requests' being conducted on their

server resulting in our requests being blocked. We fixed this issue by adding system sleep to reduce the chance of this issue. However, this may not an adequate solution for larger chemical lists. Other possibilities include inputting the maximum number of chemicals in the search box for each search (200), rather than imputing chemicals and running searches one at a time. We maximized the speed of downloading data by completing the loop with a CSV link that could be downloaded separately. Overall, ADME retrieval of 166 chemicals took roughly 20 minutes to run.

Physiochemical and ADME Properties

We found a significant difference in mean molecular weight between at least two chemical classes in sagebrush species ($F_{2,163} = 39.80$, $p < 0.0001$, Figure 1.2). The molecular weight was significantly higher for phenolics than monoterpenes ($p < 0.0001$, 95% C.I. = [90.36 – 195.04]) and higher for sesquiterpenes than monoterpenes ($p < 0.0001$, 95% C.I. = [85.17 – 156.91]) but did not differ between phenolics and sesquiterpenes ($p = 0.593$, 95% C.I. = [-30.85 – 74.17]). We also found a significant difference in lipophilicity between at least two chemical classes ($F_{2,163} = 16.68$, $p < 0.0001$, Figure 1.3). Lipophilicity was significantly lower for phenolics than sesquiterpenes ($p < 0.0001$, 95% C.I. = [0.44 – 1.08]) and lower than monoterpenes ($p < 0.0001$, 95% C.I. = [0.36 – 1.00]) but did not differ between monoterpenes and sesquiterpenes ($p = 0.65$, 95% C.I. = [-0.14 – 0.30]).

Chemical Ecology Predictions

Chemical classes did not differ in their predicted GI absorption (X^2 (2, $N = 166$) = 0.756, $p = 0.685$, Figure 1.4 a) with 74% of monoterpenes, 77% of phenolics, and 69% of sesquiterpenes having high GI absorption. Only a small proportion of sesquiterpenes

(17%) and phenolics (5%) were predicted to be P-gp substrates compared to none of the monoterpenes (X^2 (2, N = 166) = 14.624, p = 0.0007) (Figure 1.4 b). However, monoterpenes and sesquiterpenes were less likely to be CYP inhibitors (3% and 7% predicted to inhibit CYP1A2, respectively) than phenolics (X^2 (2, N = 166) = 74.202, p = <0.0001) where 72.73% of phenolics were predicted to inhibit CYP1A2 (Figure 1.4 c). Monoterpenes that are induced or avoided by foraging herbivores were more likely to have lower predicted GI absorption than monoterpenes that were not induced or avoided (X^2 (1, N = 24) = 9.882, p = 0.0017, Figure 1.5 a). Neither being a P-gp substrate (X^2 (1, N = 24) = 0, p = 0.0 or null) nor inhibiting CYP1A2 (X^2 (1, N = 24) = 1.043, p = 0.307) differed based on whether monoterpenes were observed to be induced or avoided by herbivores (Figure 1.5 b and c).

Discussion

Accurately and effectively retrieving large amounts of predicted PK data using automation would be incredibly beneficial for chemical ecologists. Here, we demonstrated how using Python and associated packages can automate the process of mining predicted PK data for chemicals found in sagebrush that might explain interactions with wild herbivores. Chemicals found in sagebrush include monoterpenes and sesquiterpenes as well as phenolics such as flavonoids and coumarins among other chemicals. Moreover, the complexity of these chemicals can be further increased through interactions with enzymes in plants, animals, and microbes, which involve multiple and often simultaneous biosynthesis and metabolism of PSMs. Focusing on this chemically diverse sagebrush system offers an example of obtaining PK data for a diverse set of chemicals. Moreover, we provide examples of how the chemical structures and predicted

ADME parameters generated from our automated pipeline can be used to test chemical ecology hypotheses that may explain avoidance or selection of PSMs by herbivores. We now discuss limitation and opportunities for further automation of PK and how our automated pipeline could advance our understanding of the fate of chemicals in a variety of systems.

Testing Accuracy

Testing for accuracy of this pipeline was done manually by undergraduate students enrolled in a Lab-based Undergraduate Research Experience (LURE, See Chapter 2). Students manually conducted the steps of the pipeline using the same chemical lists, including monoterpenes, phenolics, and sesquiterpenes (Turi et al., 2014) and compared the manually mined PK output they obtained with the output of the automated pipeline. Students discovered that 53 chemicals were not found on PubChem and allowed us to correct the spelling and annotation to include 14 of these missing chemicals into our revised pipeline.

Further Automation

Our initial effort to automate PK resulted in several insights that could benefit further automation. First, we identified the importance of the original data sheet for getting accurate output for the desired data input. Even after automation was achieved, each output step revealed issues that may require manual assessment (e.g., using an undergraduate workforce in the classroom, See Chapter 2) and computational revision (Table 1.1). Based on our experience, we recommend adding quality control steps to check for single chemical names and proper spelling of chemicals in each cell. This quality control step could be done by viewing the “undoundcid” list and manually

determining possible reasons for the failure to obtain CID information from PubChem. In addition, extracting chemical names directly from original publications could improve the amount of data and accuracy of data that can be produced from this pipeline. We included chemical names from a single review paper (Turi et al. 2014). However, the list of chemicals could be expanded by integrating other data mining packages such as G2PMineR package (Wojahn et al., 2021) to search and obtain a more diverse set of peer-reviewed scientific research papers. These papers can be downloaded as PDF files and specifically designed packages (e.g., requests Chandra & Varanasi, 2015) can be used to extract tables from PDF files to run through our pipeline.

To avoid large chemical lists from being blocked from the SwissADME database mid-search, we recommend using system sleep on Python to prevent lockout or using batches of 200 chemicals to run the search. However, we encourage collaboration with administrators of open access databases to facilitate knowledge transfer between users and generators of chemical structures and PK data. We envision that the combination of extracting chemical names from all available publications, using system sleep features, and collaboration with the SwissADME database administrators could be combined to continually generate diverse lists of chemicals with associated ADME parameters. In addition, we recommend leveraging undergraduate courses as a sustainable quality control step to continuously validate structures and mined ADME data and to propose and test new hypotheses while also providing students with authentic research experience and training that integrates chemistry, pharmacology, and computing (see Chapter 2).

Researchers replicating our pipeline should consider limitations and unpredictable issues that could affect the functionality of the database and the automation of the

pipeline created. Use of a chemical formula rather than the canonical SMILES notation of the chemical may result in inaccurate PK output because multiple chemicals may have the same chemical formula or conformational isomers (Tang et al., 2020). For example, 21 of the chemicals found in Turi et al. (2014) had the chemical formula C10H16O. In addition, as a molecule gets larger, it is harder to identify the chemical properly which would further reduce the accuracy of PK output from a chemical formula. If databases make major changes to their JavaScript code, the created pipeline would also have to be updated to match changes. One solution is for databases to use an Application Programming Interface (API) to share data and execute predefined functionality. Currently, SwissADME does not have an API which restricts access of the database is through their website. Having an API or an associated package in the SwissADME database would increase capacity for researchers to obtain important information quickly and accurately for a large number of input chemicals.

Advance Understanding of PK

Our pipeline increased the capacity of chemical ecologists, pharmacologists, and toxicologists to quickly, effectively, and efficiently mine predicted PK values. It is of major importance for scientists to understand PK to advance our understanding of plant-herbivore-microbial interactions, drug-drug interactions, and develop treatments for patients in need. In our case study, we focused on PSMs in sagebrush identified from the literature to mine predicted PK information. However, our automated approach could be used to obtain ADME parameters beyond parent chemicals by including structures of new metabolites detected using analytical chemistry (e.g., mass spectrometry [Liu et al., 2019] and NMR [Kim et al., 2011; Seger & Sturm, 2022]) that are generated from

microbial cultures (Kohl et al., 2018) or from metabolic stability assays (e.g., cytochrome P450s from microsomes, Di et al., 2008) using commercially available microbes and enzymes or those isolated from non-model organisms (e.g., wild herbivores) that are inoculated with parent chemicals. Each of these sources offers diverse metabolites, many of which not yet identified, that provide endless possibilities for discovering diverse structures that can aid in understanding and managing the health of humans, domestic species, ad non-model organisms.

Understanding the journey of a chemical from the time of consumption to excretion is very important and can be predicted based on physiochemical properties. If the molecular weight is low and lipophilicity is high, it is likely the chemical will have high GI absorption. Once the chemical is absorbed it may be a substrate or an inhibitor of efflux transporter proteins like P-gp. If the chemical is a substrate of P-gp, then it will be effluxed out of the cell back into the lumen of the intestine where it may be metabolized by microbes associated with the herbivore intestines (Ahmed Juvale et al., 2022, Duda-Chodak et al., 2015). If the chemical is not a substrate of P-gp, then it will be absorbed into the enterocyte and distributed to the liver where it can interact (substrates or inhibitors) with host metabolizing enzymes such as CYPs. If the chemical is a substrate of CYPs it will get metabolized, if it is an inhibitor, it can inhibit the metabolism of any co-occurring chemicals (Zhou, 2008). After absorption, the distribution and excretion of the chemical is based on interactions with both efflux transporters and metabolizing enzymes in tissues (Fan et al., 2010). Although there are many mechanisms within the body of the herbivore that may affect the ADME of the chemical, we focused on the P-gp

transporter substrates (Ahmed Juvale et al., 2022) and CYP1A2 inhibition (Gunes & Dahl, 2008).

Specific to sagebrush, we used our pipeline to identify interesting patterns in PSMs that may explain some of the plant-herbivore interactions observed in this system. First, our pipeline demonstrated that monoterpenes were smaller than other chemical classes in sagebrush and had relatively high lipophilicity ($logP_{n\text{-}octanol/water}$, scale -5 to 5, if $logP_{o/w}$ value is 2 that means the drug is 100 times more likely to be in n-octanol than water) with limited variation among chemicals. These traits indicate that monoterpenes are more likely to be absorbed, distributed into tissue, and be bioactive (van de Waterbeemd et al., 2001). In contrast, SwissADME predicted that monoterpenes, despite smaller size and higher lipophilicity, did not differ in GI absorption compared to phenolics and sesquiterpenes. However, there was evidence that monoterpenes that were induced or avoided by herbivores were less likely to have high GI absorption. This suggests that the site of action for monoterpenes induced and avoided may be in the intestine of the consumer. PSMs that have low GI absorption remain in the GI tract and are not absorbed into the bloodstream, but instead are metabolized in the GI tract by microbes or are excreted in the feces of the herbivore. Regulated absorption of monoterpenes observed in sage-grouse (Frye, 2012) would reduce systemic exposure of these toxins to wildlife which would minimize harmful effects and explain the relatively high tolerance sage-grouse have for consuming monoterpenes. The structure of monoterpenes suggest GI absorption in most cases should be high, but sagebrush specialist herbivores have mechanisms that actively reduce absorption of toxic monoterpenes (Forbey et al., 2011) or allow them to resist inhibition of digestive

enzymes by monoterpenes (Kohl et al., 2015). Interestingly, of the seven monoterpenes that were studied previously (Frye, 2012), three of which (methacrolein, 1,8-cineole, and camphor) had high GI absorption and the other four (α-pinene, β-pinene, camphene, and cymene) had low GI absorption. Camphor and 1,8-cineole are known to be more slowly metabolized by rabbits when compared with other monoterpenes (Nobler et al., 2019). All seven of these monoterpenes had high lipophilicity (between 2.12 and 2.65, except for methacrolein which had 1.25), none of these chemicals were CYP1A2 inhibitors. However, α -pinene, β-pinene, camphene, and cymene were all inhibitors of different CYP enzymes (α -pinene inhibited CYP2C9; β-pinene inhibited CYP2C9; camphene inhibited CYP2C9; and cymene inhibited CYP2D6). Next steps would be to test the PK predicted inhibition of each monoterpene using *in vitro* metabolic stability assays that can be done using specific commercially available CYP isoforms or done using microsomes isolated from herbivores of interest.

Another interesting finding was that phenolics were most likely to be CYP1A2 inhibitors and sesquiterpenes were most likely to be P-gp substrates along with one chemical from the phenolics chemical class (Figure 1.4). Based on this pattern, we predict that phenolics have relatively high potential to influence the PK of co-occurring chemicals through CYP inhibition and that sesquiterpenes may be the least bioavailable because as P-gp substrates, they have the greatest potential to be effluxed out of cells and remain in the intestine. We caution that *in silico* predictions do not always translate to *in vivo* outcomes (Jain et al., 2021) and using *in silico* models to differentiate between P-gp substrate and inhibitors remains complicated (Chen et al., 2018). Proposed next steps for this research would be to test patterns of *in vitro* intestinal activity of P-gp in avian and

mammalian herbivores (Green et al., 2005). In our case study, we used a subset of chemicals in *Artemisia*. However, we have established new capacity to make predictions such as those between sagebrush and other *Artemisia* species with known PK and therapeutic properties (Bora & Sharma, 2011) that can inform more labor intensive and expensive *in vitro* and *in vivo* assays. Finally, results demonstrate how mined PK data provide insight into existing chemical interactions between plants and herbivores and can be used to target specific chemical-enzyme combinations in future *in vitro* experiments.

Conclusion

The purpose of this case study was to automate the process of mining predicted PK of complex chemicals found in the genus *Artemisia*. Based on this case study, it can be concluded that automating this multi-step process can save time and allow researchers a more reproducible and transparent approach for PK data retrieval, which can also allow reproducing this analysis when new compounds are identified. Our reproducible pipeline has improved the ability of scientists to determine the PK of chemicals for large scale studies by offering an easy and fast tool for future analysis. Using our pipeline, we have expanded capacity to predict PK data for diverse classes of chemicals that play a role in plant-herbivore interactions. This pipeline will help scientists understand and predict PK of a large number of chemicals which can not only help us understand concentrations of potentially toxic PSMs that explain the foraging ecology of herbivores (Williams et al., 2022), but also help us identify natural products in sagebrush that could contribute to biomedical research (Atanasov et al., 2021; Newman & Cragg, 2020; Seo et al., 2020). Future work on this project would include creating an API or a Python package for this pipeline to make it more accessible to all scientists and creating a database of ADME

data for molecules found in other species within the genus *Artemisia* and in other plant taxa that are consumed by herbivores. These future steps are particularly useful for chemical ecologists who work with a broad range of compounds which may include complex unknown chemicals. Our pipeline requires only that scientists can provide structures, not the identity, of compounds to mine ADME data. Our pipeline overcomes existing bottlenecks of databases which may limit the number of chemicals that may be searched at one time. Our automation of the manual multi-step retrieval process allows for a larger number of compounds to be efficiently and effectively searched at one time.

The ability to predict exposure of metabolites is essential to understanding and managing the health of humans, domestic species, and non-model organisms. Sagebrush provides an excellent case study to demonstrate the impact of these predictions because it produces chemicals that are used as drugs (Nagy & Tengerdy, 1967), the plant can be used as food for livestock (Burritt et al., 2000), and the chemical quality of sagebrush influences the foraging behavior of free-ranging herbivores (Willms et al., 1980) including those of conservation concern (e.g., sage-grouse (Welch et al., 1988) and pygmy rabbits (Jimenez et al., 2020). The ADME parameters generated from plant chemicals could also be used to optimize dosing of drugs in humans (Lucas et al., 2019), select plants for reseeding that are most palatable to livestock (Copeland et al., 2021), and interpret chemical resistance in dietary specialist herbivores and dietary avoidance by generalist herbivores (Boyle et al., 1999, Nobler et al., 2019).

References

Ahmad, A. M. (2007). Recent advances in pharmacokinetic modeling. *Biopharmaceutics & Drug Disposition*, *28*(3), 135–143. https://doi.org/10.1002/bdd.540

Ahmed Juvale, I. I., Abdul Hamid, A. A., Abd Halim, K. B., & Che Has, A. T. (2022). P-glycoprotein: New insights into structure, physiological function, regulation and alterations in disease. *Heliyon*, *8*(6), e09777. https://doi.org/10.1016/j.heliyon.2022.e09777

Alavijeh, M. S., Chishty, M., Qaiser, M. Z., & Palmer, A. M. (2005). Drug metabolism and pharmacokinetics, the blood-brain barrier, and central nervous system drug discovery. *NeuroRx*, *2*(4), 554–571.

Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., & Supuran, C. T. (2021). Natural products in drug discovery: Advances and opportunities. *Nature Reviews Drug Discovery*, *20*(3), Article 3. https://doi.org/10.1038/s41573-020-00114-z

Bagale, S. (2022). An overview on polychlorinated biphenyls pollutants. *Journal of Bioremediation and Biodegradation*, *13*(8).

Bora, K. S., & Sharma, A. (2011). The genus *Artemisia*: A comprehensive review. *Pharmaceutical Biology*, *49*(1), 101–109. https://doi.org/10.3109/13880209.2010.497815

Boyle, R., McLean, S., Foley, W. J., & Davies, N. W. (1999). Comparative metabolism of dietary terpene, p-cymene, in generalist and specialist folivorous marsupials. *Journal of Chemical Ecology*, *25*(9), 2109–2126.

Burritt, E. A., Banner, R. E., & Provenza, F. D. (2000). Sagebrush ingestion by lambs: Effects of experience and macronutrients. *Journal of Range Management*, *53*(1), 91. https://doi.org/10.2307/4003398

Carvalhais, L. C., Dennis, P. G., Fan, B., Fedoseyenko, D., Kierul, K., Becker, A., von Wiren, N., & Borriss, R. (2013). Linking plant nutritional status to plant-microbe interactions. *PLoS ONE*, *8*(7), Article 7. https://doi.org/10.1371/journal.pone.0068555

Chandra, R. V., & Varanasi, B. S. (2015). *Python requests essentials*. Packt Publishing.

Chen, C., Lee, M.-H., Weng, C.-F., & Leong, M. K. (2018). Theoretical prediction of the complex p-glycoprotein substrate efflux based on the novel hierarchical support vector regression scheme. *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry*, *23*(7), 1–25. https://doi.org/10.3390/molecules23071820

Copeland, S. M., Davies, K. W., Boyd, C. S., & Bates, J. D. (2021). Recovery of the herbaceous component of degraded sagebrush steppe is unimpeded by 75 years of moderate cattle grazing. *Ecosphere*, *12*(3), e03445. https://doi.org/10.1002/ecs2.3445

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/srep42717

Di, L., Kerns, E. H., Ma, X. J., Huang, Y., & Carter, G. T. (2008). Applications of high throughput microsomal stability assay in drug discovery. *Combinatorial Chemistry & High Throughput Screening*, *11*(6), 469–476. https://doi.org/10.2174/138620708784911429

Díaz-Navarro, M., Bolívar, P., Andrés, M. F., Gómez-Muñoz, M. T., Martínez-Díaz, R. A., Valcárcel, F., García-París, M., Bautista, L. M., & González-Coloma, A. (2021). Antiparasitic effects of potentially toxic beetles *(Tenebrionidae and Meloidae)* from steppe zones. *Toxins*, *13*(7), Article 7. https://doi.org/10.3390/toxins13070489

Duda-Chodak, A., Tarko, T., Satora, P., & Sroka, P. (2015). Interaction of dietary compounds, especially polyphenols, with the intestinal microbiota: A review. *European Journal of Nutrition*, *54*(3), 325–341. https://doi.org/10.1007/s00394-015-0852-y

Dyer, L. A., Philbin, C. S., Ochsenrider, K. M., Richards, L. A., Massad, T. J., Smilanich, A. M., Forister, M. L., Parchman, T. L., Galland, L. M., Hurtado, P. J., Espeset, A. E., Glassmire, A. E., Harrison, J. G., Mo, C., Yoon, S., Pardikes, N. A., Muchoney, N. D., Jahner, J. P., Slinn, H. L., … Jeffrey, C. S. (2018). Modern approaches to study plant–insect interactions in chemical ecology. *Nature Reviews Chemistry*, *2*(6), Article 6. https://doi.org/10.1038/s41570-018-0009-7

Eddershaw, P. J., Beresford, A. P., & Bayliss, M. K. (2000). ADME/PK as part of a rational approach to drug discovery. *Drug Discovery Today*, *5*(9), 409–414. https://doi.org/10.1016/S1359-6446(00)01540-3

Ekrami, A., Masoudian, N., Mahmoodzadeh, H., Roodi, B., & Ebadi, M. (2022). Effects of altitude on some physiological characteristics of sagebrush in Khorasan Province, Iran. *Journal of Chemical Health Risks*, *12*. https://doi.org/10.22034/jchr.2022.580687.1006

Fan, J., Chen, S., C.Y. Chow, E., & Sandy Pang, K. (2010). PBPK modeling of intestinal and liver enzymes and transporters in drug absorption and sequential metabolism. *Current Drug Metabolism*, *11*(9), 743–761. https://doi.org/10.2174/138920010794328931

Forbey, J. S., Dearing, M. D., Gross, E. M., Orians, C. M., Sotka, E. E., & Foley, W. J. (2013). A pharm-ecological perspective of terrestrial and aquatic plant-herbivore interactions. *Journal of Chemical Ecology*, *39*(4), 465–480. https://doi.org/10.1007/s10886-013-0267-2

Forbey, J., Frye, G., & Pu, X. (2011). Toxic scat: A mechanism to prevent overdosing on plant chemicals by grouse. *Grouse News*, *42*. https://scholarworks.boisestate.edu/bio_facpubs/263

Freeland, W. J., & Janzen, D. H. (1974). Strategies in herbivory by mammals: The role of plant secondary compounds. *The American Naturalist*, *108*(961), 269–289. https://doi.org/10.1086/282907

Frye, G. G. (2012). *Phytochemical ecology of an avian herbivore, the greater sage-grouse: Implications for behavior, physiology, and conservation*. Boise State University Theses and Dissertations.

Gfeller, D., Grosdidier, A., Wirth, M., Daina, A., Michielin, O., & Zoete, V. (2014). SwissTargetPrediction: A web server for target prediction of bioactive small molecules. *Nucleic Acids Research*, *42*(Web Server issue), W32–W38. https://doi.org/10.1093/nar/gku293

Godfray, H. C. J., Stephens, A. E. A., Jepson, P. D., Jobling, S., Johnson, A. C., Matthiessen, P., Sumpter, J. P., Tyler, C. R., & McLean, A. R. (2019). A restatement of the natural science evidence base on the effects of endocrine disrupting chemicals on wildlife. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1897), 1–11. https://doi.org/10.1098/rspb.2018.2416

Green, A. K., Barnes, D. M., & Karasov, W. H. (2005). A new method to measure intestinal activity of P-glycoprotein in avian and mammalian species. *Journal of Comparative Physiology. B, Biochemical, Systemic, and Environmental Physiology*, *175*(1), 57–66. https://doi.org/10.1007/s00360-004-0462-0

Gunes, A., & Dahl, M.-L. (2008). Variation in CYP1A2 activity and its clinical implications: Influence of environmental factors and genetic polymorphisms. *Pharmacogenomics*, *9*(5), 625–637. https://doi.org/10.2217/14622416.9.5.625

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), Article 7825. https://doi.org/10.1038/s41586-020-2649-2

Huang, Y., Cao, S., Zhang, Q., Zhang, H., Fan, Y., Qiu, F., & Kang, N. (2018). Biological and pharmacological effects of hexahydrocurcumin, a metabolite of curcumin. *Archives of Biochemistry and Biophysics*, *646*, 31–37. https://doi.org/10.1016/j.abb.2018.03.030

Jain, S., Norinder, U., Escher, S. E., & Zdrazil, B. (2021). Combining in vivo data with in silico predictions for modeling hepatic steatosis by using stratified bagging and conformal prediction. *Chemical Research in Toxicology*, *34*(2), 656–668. https://doi.org/10.1021/acs.chemrestox.0c00511

Jimenez, A., Schmalz, J., Wright, M. N., & Skopec, M. M. (2020). Sagebrush characteristics influencing foraging patterns of pygmy rabbits. *The Journal of Wildlife Management*, *84*(7), 1306–1314. https://doi.org/10.1002/jwmg.21923

Karunanithi, P. S., & Zerbe, P. (2019). Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. *Frontiers in Plant Science*, *10*. https://www.frontiersin.org/articles/10.3389/fpls.2019.01166

Kato, H. (2020). Computational prediction of cytochrome P450 inhibition and induction. *Drug Metabolism and Pharmacokinetics*, *35*(1), 30–44. https://doi.org/10.1016/j.dmpk.2019.11.006

Kay, C. D. (2006). Aspects of anthocyanin absorption, metabolism and pharmacokinetics in humans. *Nutrition Research Reviews*, *19*(1), 137–146. https://doi.org/10.1079/NRR2005116

Kelley, B. D., Appelt, J. M., & Appelt, G. D. (1992). *Artemisia tridentata* (basin sagebrush) in the southwestern United States of America: Medicinal uses and pharmacologic implications. *International Journal of the Addictions*, *27*(3), 347–366. https://doi.org/10.3109/10826089209068747

Khalifa, S. A. M., Elias, N., Farag, M. A., Chen, L., Saeed, A., Hegazy, M.-E. F., Moustafa, M. S., Abd El-Wahed, A., Al-Mousawi, S. M., Musharraf, S. G., Chang, F.-R., Iwasaki, A., Suenaga, K., Alajlani, M., Göransson, U., & El-Seedi, H. R. (2019). Marine natural products: A source of novel anticancer drugs. *Marine Drugs*, *17*(9), Article 9. https://doi.org/10.3390/md17090491

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Research*, *49*(D1), D1388–D1395. https://doi.org/10.1093/nar/gkaa971

Kohl, K. D., Oakeson, K. F., Orr, T. J., Miller, A. W., Forbey, J. S., Phillips, C. D., Dale, C., Weiss, R. B., & Dearing, M. D. (2018). Metagenomic sequencing provides insights into microbial detoxification in the guts of small mammalian herbivores (*Neotoma* spp.). *FEMS Microbiology Ecology*, *94*(12), 1–10. https://doi.org/10.1093/femsec/fiy184

Kohl, K. D., Pitman, E., Robb, B. C., Connelly, J. W., Dearing, M. D., & Forbey, J. S. (2015). Monoterpenes as inhibitors of digestive enzymes and counter-adaptations in a specialist avian herbivore. *Journal of Comparative Physiology B*, *185*(4), 425–434. https://doi.org/10.1007/s00360-015-0890-z

Kohl, K., & Dearing, M. (2017). Intestinal lymphatic transport: An overlooked pathway for understanding absorption of plant secondary compounds in vertebrate herbivores. *Journal of Chemical Ecology*, *43*(3), Article 3. https://doi.org/10.1007/s10886-017-0828-x

Krause, T., Moenning, J.-L., Lamp, J., Maul, R., Schenkel, H., Fürst, P., Pieper, R., & Numata, J. (2022). Transfer of polychlorinated dibenzo-p-dioxins and dibenzofurans (PCDD/Fs) and polychlorinated biphenyls (PCBs) from oral exposure into cow's milk – Part I: State of knowledge and uncertainties. *Nutrition Research Reviews*, 1–23. https://doi.org/10.1017/S0954422422000178

Lautié, E., Russo, O., Ducrot, P., & Boutin, J. A. (2020). Unraveling plant natural chemical diversity for drug discovery purposes. *Frontiers in Pharmacology*, *11*. https://www.frontiersin.org/articles/10.3389/fphar.2020.00397

Leven, C., Padelli, M., Carré, J.-L., Bellissant, E., & Misery, L. (2019). Immune checkpoint inhibitors in melanoma: A review of pharmacokinetics and exposure–response relationships. *Clinical Pharmacokinetics*, *58*(11), 1393–1405. https://doi.org/10.1007/s40262-019-00789-7

Li, J., T. Baldwin, I., & Li, D. (2022). Harmonizing biosynthesis with post-ingestive modifications to understand the ecological functions of plant natural products. *Natural Product Reports*, *39*(7), 1383–1392. https://doi.org/10.1039/D2NP00019A

Liu, X., Zhou, L., Shi, X., & Xu, G. (2019). New advances in analytical methods for mass spectrometry-based large-scale metabolomics study. *TrAC Trends in Analytical Chemistry*, *121*, 1–11. https://doi.org/10.1016/j.trac.2019.115665

Lombardo, F., Desai, P. V., Arimoto, R., Desino, K. E., Fischer, H., Keefer, C. E., Petersson, C., Winiwarter, S., & Broccatelli, F. (2017). In silico absorption, distribution, metabolism, excretion, and pharmacokinetics (ADME-PK): Utility and best practices. an industry perspective from the international consortium for innovation through quality in pharmaceutical development. *Journal of Medicinal Chemistry*, *60*(22), 9097–9113. https://doi.org/10.1021/acs.jmedchem.7b00487

Lucas, A. J., Sproston, J. L., Barton, P., & Riley, R. J. (2019). Estimating human ADME properties, pharmacokinetic parameters and likely clinical dose in drug discovery. *Expert Opinion on Drug Discovery*, *14*(12), 1313–1327. https://doi.org/10.1080/17460441.2019.1660642

Lutz, R. J., Dedrick, R. L., Tuey, D., Sipes, I. G., Anderson, M. W., & Matthews, H. B. (1984). Comparison of the pharmacokinetics of several polychlorinated biphenyls in mouse, rat, dog, and monkey by means of a physiological pharmacokinetic model. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *12*(5), 527–535.

Ma, S., McGann, M., & Enyedy, I. J. (2021). The influence of calculated physicochemical properties of compounds on their ADMET profiles. *Bioorganic & Medicinal Chemistry Letters*, *36*, 127825. https://doi.org/10.1016/j.bmcl.2021.127825

Massing, L. T., Suemitsu, C., Sarrazin, S. L. F., Tremea, A., Maia, J. G. S., & Mourão, R. H. V. (2021). Oils variability of *Pectis elongata* in the Amazon and an overview of the neotropical *Pectis* species. *European Journal of Medicinal Plants*, *32*(1), 37–45. https://doi.org/10.9734/ejmp/2021/v32i130362

Maurya, P. K., Malik, D. S., Yadav, K. K., Kumar, A., Kumar, S., & Kamyab, H. (2019). Bioaccumulation and potential sources of heavy metal contamination in fish species in River Ganga basin: Possible human health risks evaluation. *Toxicology Reports*, *6*, 472–481. https://doi.org/10.1016/j.toxrep.2019.05.012

Mazid, M., Khan, T. A., & Mohammad, F. (2011). Role of secondary metabolites in defense mechanisms of plants. *Biology and Medicine*, *3*(2), 232–249.

Mazoit, J. X., Butscher, K., & Samii, K. (2007). Morphine in postoperative patients: Pharmacokinetics and pharmacodynamics of metabolites. *Anesthesia & Analgesia*, *105*(1), Article 1. https://doi.org/10.1213/01.ane.0000265557.73688.32

Mazorra-Alonso, M., Tomás, G., & Soler, J. J. (2021). Microbially mediated chemical ecology of animals: A review of its role in conspecific communication, parasitism and predation. *Biology*, *10*(4), Article 4. https://doi.org/10.3390/biology10040274

McKinney, W. (2010). *Data structures for statistical computing in Python*. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

McLean, S., Boyle, R. R., Brandon, S., Davies, N. W., & Sorensen, J. S. (2007). Pharmacokinetics of 1,8-cineole, a dietary toxin, in the brushtail possum *(Trichosurus vulpecula)*: Significance for feeding. *Xenobiotica*, *37*(9), 903–922. https://doi.org/10.1080/00498250701570277

Mushtaq, S., Abbasi, B. H., Uzair, B., & Abbasi, R. (2018). Natural products as reservoirs of novel therapeutic agents. *EXCLI Journal*, *17*, 420–451. https://doi.org/10.17179/excli2018-1174

Nagy, J. G., & Tengerdy, R. P. (1967). Antibacterial action of essential oils of *Artemisia* as an ecological factor. *Applied Microbiology*, *15*(4), 819–821. https://doi.org/10.1128/am.15.4.819-821.1967

Newman, D. J., & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products*, *83*(3), 770–803. https://doi.org/10.1021/acs.jnatprod.9b01285

Nobler, J. D., Camp, M. J., Crowell, M. M., Shipley, L. A., Dadabay, C., Rachlow, J. L., James, L., & Forbey, J. S. (2019). Preferences of specialist and generalist mammalian herbivores for mixtures versus individual plant secondary metabolites. *Journal of Chemical Ecology*, *45*(1), Article 1. https://doi.org/10.1007/s10886-018-1030-5

Oh, K. P., Aldridge, C. L., Forbey, J. S., Dadabay, C. Y., & Oyler-McCance, S. J. (2019). Conservation genomics in the sagebrush sea: Population divergence, demographic history, and local adaptation in sage-grouse (*Centrocercus* spp.). *Genome Biology and Evolution*, *11*(7), 2023–2034. https://doi.org/10.1093/gbe/evz112

Parham, F. M., Kohn, M. C., Matthews, H. B., Derosa, C., & Portier, C. J. (1997). Using structural information to create physiologically based pharmacokinetic models for all polychlorinated biphenyls. *Toxicology and Applied Pharmacology*, *144*(2), 340–347. https://doi.org/10.1006/taap.1997.8139

Pathania, S., Bhatia, R., Baldi, A., Singh, R., & Rawal, R. K. (2018). Drug metabolizing enzymes and their inhibitors' role in cancer resistance. *Biomedicine & Pharmacotherapy*, *105*, 53–65. https://doi.org/10.1016/j.biopha.2018.05.117

Radeva-llieva, M., Stoeva, S., Hvarchanova, N., Zhelev, I., & Georgiev, K. D. (2022). Influence of methylxanthines isolated from *Bancha* green tea on the pharmacokinetics of sildenafil in rats. *DARU Journal of Pharmaceutical Sciences*, *30*(1), 75–84. https://doi.org/10.1007/s40199-022-00433-z

Rajaganapathy, V., Xavier, F., Sreekumar, D., & Mandal, P. K. (2011). Heavy metal contamination in soil, water and fodder and their presence in livestock and products: A review. *Journal of Environmental Science and Technology*, *4*(3), 234–249.

Sánchez, E., García, S., & Heredia, N. (2010). Extracts of edible and medicinal plants damage membranes of *Vibrio cholerae*. *Applied and Environmental Microbiology*, *76*(20), 6888–6894. https://doi.org/10.1128/AEM.03052-09

Seger, C., & Sturm, S. (2022). NMR-based chromatography readouts: Indispensable tools to "translate" analytical features into molecular structures. *Cells*, *11*(21), Article 21. https://doi.org/10.3390/cells11213526

Seo, M., Shin, H. K., Myung, Y., Hwang, S., & No, K. T. (2020). Development of natural compound molecular fingerprint (NC-MFP) with the dictionary of natural products (DNP) for natural product-based drug development. *Journal of Cheminformatics*, *12*(6), 1–17. https://doi.org/10.1186/s13321-020-0410-3

Shafranovich, Y. (2005). *Common format and MIME type for comma-separated values (CSV) files* (Request for Comments RFC 4180). Internet Engineering Task Force. https://doi.org/10.17487/RFC4180

Sharifi-Rad, J., Ozleyen, A., Boyunegmez Tumer, T., Oluwaseun Adetunji, C., El Omari, N., Balahbib, A., Taheri, Y., Bouyahya, A., Martorell, M., Martins, N., & C. Cho, W. (2019). Natural products and synthetic analogs as a source of antitumor drugs. *Biomolecules*, *9*(11), Article 11. https://doi.org/10.3390/biom9110679

Sorensen, J. S., McLister, J. D., & Dearing, M. D. (2005). Novel plant secondary metabolites impact dietary specialists more than generalists (*Neotoma* spp.). *Ecology*, *86*(1), 140–154.

Sorensen, J. S., Turnbull, C. A., & Dearing, M. D. (2004). A specialist herbivore *(Neotoma stephensi)* absorbs fewer plant toxins than does a generalist *(Neotoma albigula)*. *Physiological & Biochemical Zoology*, *77*(1), 139–148. https://doi.org/10.1086/378923

Staudenmaier, A. R., Shipley, L. A., Camp, M. J., Forbey, J. S., Hagerman, A. E., Brandt, A. E., & Thornton, D. H. (2022). Mule deer do more with less: Comparing their nutritional requirements and tolerances with white-tailed deer. *Journal of Mammalogy*, *103*(1), Article 1. https://doi.org/10.1093/jmammal/gyab116

Takahashi, A., & Shimada, T. (2008). Selective consumption of acorns by the Japanese wood mouse according to tannin content: A behavioral countermeasure against plant secondary metabolites. *Ecological Research*, *23*(6), 1033–1038. https://doi.org/10.1007/s11284-008-0473-5

Tang, C., Tang, Y., Ye, Y., Yan, Z., Chen, Z., Chen, L., Zhang, L., Liu, J., Shi, J., Xia, H., & Hong, W. (2020). Identifying the conformational isomers of single-molecule cyclohexane at room temperature. *Chem*, *6*(10), 2770–2781. https://doi.org/10.1016/j.chempr.2020.07.024

Turi, C. E., Shipley, P. R., & Murch, S. J. (2014). North American *Artemisia* species from the subgenus *tridentatae* (Sagebrush): A phytochemical, botanical and pharmacological review. *Phytochemistry*, *98*, 9–26. https://doi.org/10.1016/j.phytochem.2013.11.016

Ulappa, A. C., Kelsey, R. G., Frye, G. G., Rachlow, J. L., Shipley, L. A., Bond, L., Pu, X., & Forbey, J. S. (2014). Plant protein and secondary metabolites influence diet selection in a mammalian specialist herbivore. *Journal of Mammalogy*, *95*(4), Article 4. https://doi.org/10.1644/14-MAMM-A-025

van de Waterbeemd, H., Smith, D. A., & Jones, B. C. (2001). Lipophilicity in PK design: methyl, ethyl, futile. *Journal of Computer-Aided Molecular Design*, *15*(3), 273–286. https://doi.org/10.1023/A:1008192010023

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Verscheijden, L. F. M., Litjens, C. H. C., Koenderink, J. B., Mathijssen, R. H. J., Verbeek, M. M., Wildt, S. N. de, & Russel, F. G. M. (2021). Physiologically based pharmacokinetic/pharmacodynamic model for the prediction of morphine brain disposition and analgesia in adults and children. *PLOS Computational Biology*, *17*(3), e1008786. https://doi.org/10.1371/journal.pcbi.1008786

Welch, B. L., Pederson, J. C., & Rodriguez, R. L. (1988). Selection of big sagebrush by sage grouse. *The Great Basin Naturalist*, *48*(2), 274–279.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Williams, J., Siramshetty, V., Nguyễn, Đ.-T., Padilha, E. C., Kabir, Md., Yu, K.-R., Wang, A. Q., Zhao, T., Itkin, M., Shinn, P., Mathé, E. A., Xu, X., & Shah, P. (2022). Using in vitro ADME data for lead compound selection: An emphasis on PAMPA pH 5 permeability and oral bioavailability. *Bioorganic & Medicinal Chemistry*, *56*, 116588. https://doi.org/10.1016/j.bmc.2021.116588

Willms, W., Bailey, A. W., & McLean, A. (1980). Effect of burning or clipping *agropyron spicatum* in the autumn on the spring foraging behaviour of mule deer and cattle. *Journal of Applied Ecology*, *17*(1), 69–84. https://doi.org/10.2307/2402964

Wishart, D. S. (2007). Improving early drug discovery through ADME modelling. *Drugs in R & D*, *8*(6), 349–362. https://doi.org/10.2165/00126839-200708060-00003

Wojahn, J. M. A., Galla, S. J., Melton, A. E., & Buerki, S. (2021). G2PMineR: A genome to phenome literature review approach. *Genes*, *12*(2), Article 2. https://doi.org/10.3390/genes12020293

Zhou, S.-F. (2008). Drugs behave as substrates, inhibitors, and inducers of human cytochrome P450 3A4. *Current Drug Metabolism*, *9*(4), 310–322. https://doi.org/10.2174/138920008784220664

Zhu, M., Yuan, Y., Yin, H., Guo, Z., Wei, X., Qi, X., Liu, H., & Dang, Z. (2022). Environmental contamination and human exposure of polychlorinated biphenyls (PCBs) in China: A review. *Science of The Total Environment*, *805*, 150270. https://doi.org/10.1016/j.scitotenv.2021.150270

Tables

Table 1.1 Summary of possible issues in the pharmacokinetic pipeline, why the issue happens, possible solutions, and outcomes.

Issue	Why	Solution	Outcome
PubChem			
No Compound ID (CID)	-Multiple chemical names in single cell. -The chemical name was misspelled. -The chemical name was a number on PubChem (eg. our Chemical Name: Deacetoxymatricarin. Chemical Name on PubChem: 10180-88-8).	Ensure there is only one chemical name in each cell (using a Lab-Based Undergraduate Research Experience (LURE, Chapter 2). -Check chemical names are spelled correctly (using a LURE). -Possibly extracting chemical names directly from the original paper (Turi et al., 2014).	-Removal of 25% (54/209) of the chemicals of interest that did not have CIDs on PubChem (when a search was conducted using the pubchempy package). -When manually checked 33.3% (18/54) chemicals that were not found on PubChem did have entries but pubchempy was unable to produce them.
SwissADME			
To Many Requests	This error came from the SwissADME database because too many searches were coming from the same IP address at the same time.	Use system sleep to address this issue, or separate input into batches to reduce risk of being blocked.	Will allow user to run searches using a large number of chemicals.
Herf Link Download	Scraping output data from SwissADME was taking to much time.	Scrape href link from SwissADME, download all links, and concatenate data.	Output was a data sheet not quite ready for analysis.

Final Data Sheet			
Formatting	When all data sheets were concatenated there where blank spaces in the data sheet causing errors. This was because the chemicals that had no CIDs from PubChem were causing inconsistencies.	Create different lists of chemical names (Found and unfound). These lists helped us organize our data.	Produced a final data set that was accurate and provided correct ADME information for each chemical.

Figures

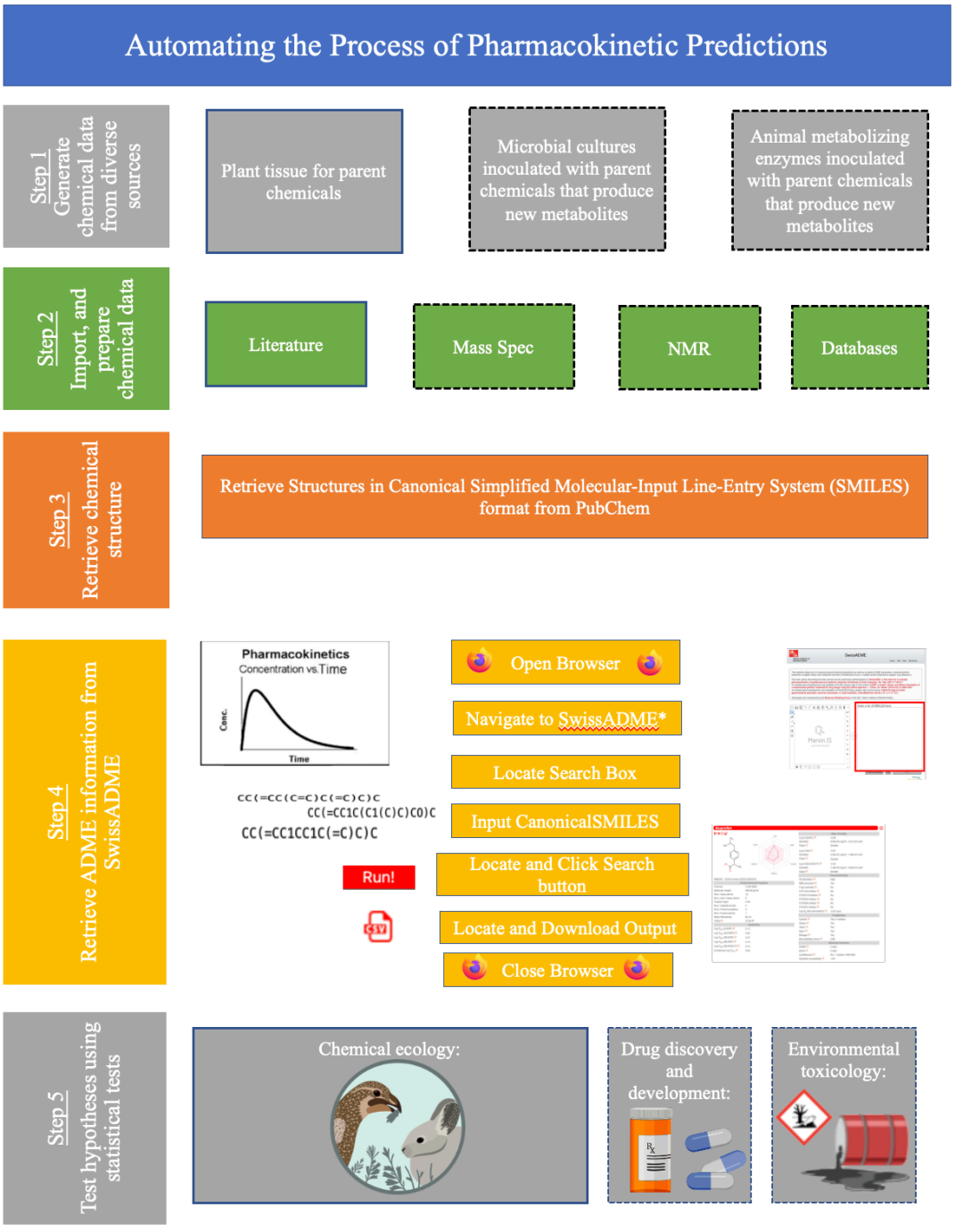


Figure 1.1 Steps of the automated pharmacokinetic pipeline. The boxes with a solid border indicate the steps that were taken for this project and the dashed borders show possible options that could be used.

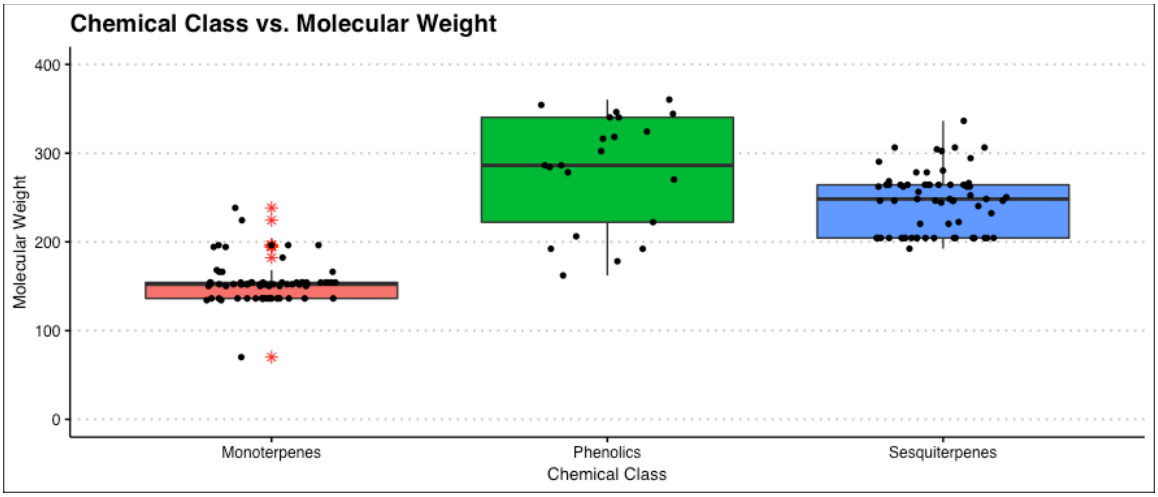


Figure 1.2 Average molecular weight (g/mol) of monoterpene (n=73), phenolic (n=22), and sesquiterpene (n=71) chemicals identified in sagebrush (*Artemisia* spp.). Outliers are signified by red asterisks. An ANOVA revealed that there was a statistically significant difference in molecular weight between at least two groups ($F_{2,163} = 39.80$, $p < 0.0001$. Tukey's HSD Test for multiple comparisons found that the mean value of molecular weight was significantly different between phenolics and monoterpenes ($p < 0.0001$, 95% C.I. = [90.36 – 195.04]). There was also a significant difference between sesquiterpenes and monoterpenes ($p < 0.0001$, 95% C.I. = [85.17 – 156.91]). There was no statistically significant difference between phenolics and sesquiterpenes ($p = 0.593$, 95% C.I. = [-30.85 – 74.17]).

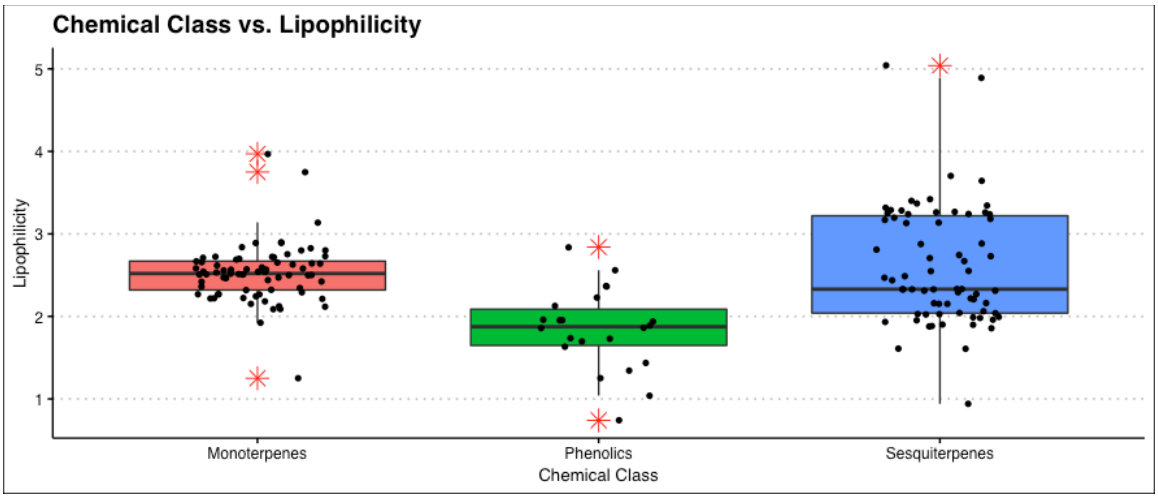


Figure 1.3 Average lipophilicity of monoterpene (n= 73), phenolic (n=22), and sesquiterpene (n=71) chemicals identified in sagebrush (*Artemisia* spp.). Outliers are signified by red asterisks. A one-way ANOVA revealed that there was a statistically significant difference in lipophilicity between at least two groups ($F_{2,163} = 16.68$, $p < 0.0001$). Tukey's HSD Test for multiple comparisons found that the mean value of lipophilicity was significantly different between phenolics and sesquiterpenes ($p < 0.0001$, 95% C.I. = [0.44 – 1.08]). There was also a significant difference between phenolics and monoterpenes ($p < 0.0001$, 95% C.I. = [0.36 – 1.00]). There was no statistically significant difference between monoterpenes and sesquiterpenes ($p = 0.65$, , 95% C.I. = [-0.14 – 0.30]).

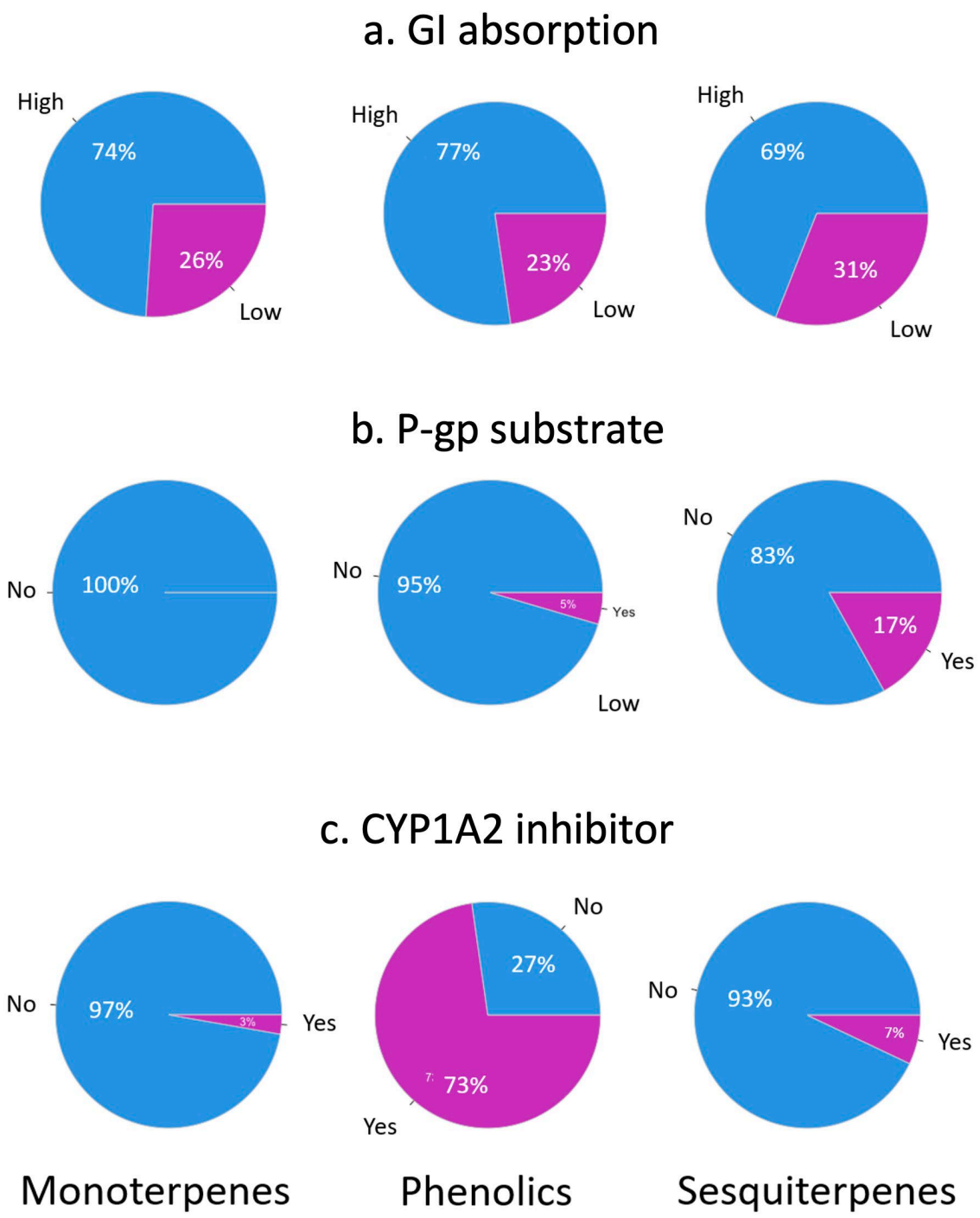


Figure 1.4 Predicted gastrointestinal (GI) absorption (a. high versus low) (X^2 (2, N = 166) = 0.756, p = 0.685); p-glycoprotein (P-gp) substrate (b, yes versus no) (X^2 (2, N = 166) = 14.624, p = 0.0007), and cytochrome P450 1A2 (CYP1A2) inhibition (c. yes versus no) (X^2 (2, N = 166) = 74.202, p = <0.0001) of monoterpenes (n=73), phenolics (n=22), and sesquiterpenes (n=71) identified in sagebrush (*Artemisia* spp.)

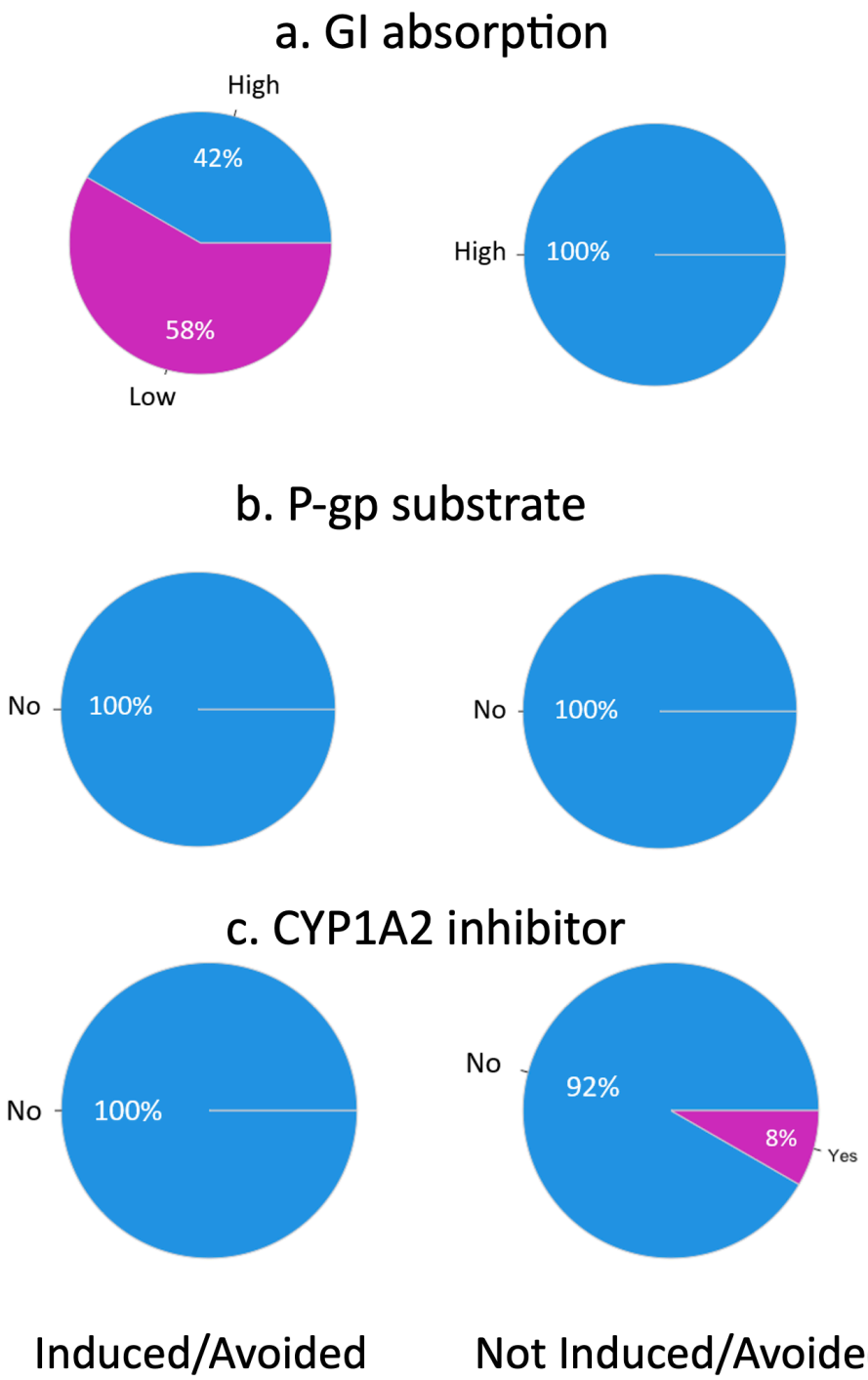


Figure 1.5 Predicted gastrointestinal (GI) absorption (a, high versus low) (X^2 (1, N = 24) = 9.882, p = 0.0017), P-gp substrate (b, yes versus no) (X^2 (1, N = 24) = 0.0, p = 0.0 or null), and cytochrome P450 1A2 (CYP1A2) inhibition (c, yes versus no) (X^2 (1, N = 24) = 1.043, p = 0.307) of monoterpenes in sagebrush (Artemisia spp.) that were observed to be induced or avoided (n=12) or not induced or avoided (n=12) by foraging herbivores.

CHAPTER TWO: LAB-BASED UNDERGRADUATE RESEARCH EXPERIENCE IN PHARMACOKINETICS

Abstract

Students pursuing an undergraduate degree in Science, Technology, Engineering, and Mathematics (STEM) greatly benefit from undergraduate research experiences. However, students may have difficulties obtaining research opportunities because of costs, time, and eligibility requirements. Similar to Course-Based Undergraduate Research Experiences (CUREs), Lab-Based Undergraduate Research Experiences (LUREs) offer equitable opportunities for students to get research experiences because they occur within lab sections of courses that they are already required to take as part of their degree program. For example, Bachelor's of Science students in Biological Sciences at Boise State University are required to take finishing foundations courses, one of which is Animal Physiology and Nutrition, where students are presented with authentic research across several LURE modules. In this course, I developed and implemented a LURE where students learned the role of pharmacokinetics (PK) in animal physiology. Students were presented with an introductory lecture about PK where they received background information about PK in pharmacology, toxicology, and chemical ecology as it relates to animal physiology. Students were also presented with a corresponding PK lab protocol where students worked both independently and collaboratively to conduct literature searches, formulate hypotheses, and present a summary of their PK findings to their peers. A survey was conducted to access our students' ability to utilize the knowledge

gained during the module and overall class. The survey reported that 100% of students increased competency in understanding why chemical structures matter in physiology and using publicly available databases to conduct research. In addition, 96% of students experiencing the PK LURE reported an increased in competency in proposing and testing hypotheses that use chemical knowledge to predict physiological research in animals. This PK LURE reinforced skills in navigating and interpreting databases, formulating directional hypotheses, and identifying and summarizing literature to support hypotheses that students generated. Through this PK LURE, 26 undergraduate students helped validate and contribute to graduate research being conducted, while also learning about graduate research and gaining authentic research experience in pharmacology that will make them more competitive for careers in veterinary medicine, pharmaceutics, and natural sciences.

Introduction

Undergraduate students in the STEM spend many hours of their education in laboratory courses immersing themselves in topics learned in the associated lecture courses. There is increasing evidence that applying their knowledge to real world experience has positive effects on learning outcomes (Beck et al., 2014). Labs that provide students authentic research experiences can develop a better understanding of topics through experimentation and hands on learning that can significantly improve associated lecture course grades as well as laboratory grades (Ing et al., 2021). Through these experiential labs, students gain knowledge, skills, and abilities in critical thinking, collaboration, analytical reasoning, and communication skills (Corwin et al., 2015) that go beyond course content and prepare them for future careers (Auchincloss et al., 2014).

The global pandemic caused courses at higher education institutions to switch to virtual instruction which left lab courses scrambling to develop lab modules students could use to engage in course content through experiential learning while remaining socially distanced (Delgado et al., 2021). Educators had to create virtual lab environments to conduct labs for the success of students (Dustman et al., 2021).

Undergraduate students interested in pursuing careers in science and research know authentic research experience are an unspoken requirement to gain employment or advanced degrees (Estrada et al., 2018). Unfortunately, not all students can join research labs. Undergraduate institutions offer a variety of research experience opportunities that may exclude some groups. Volunteer positions highly benefit privileged students who have financial support from family and have the freedom to work without monetary compensation. Paid Research Experiences for Undergraduates (REUs) summer programs offer a way for undergraduate students to obtain research experiences but often require a minimum grade point average (GPA) to be competitive due to the large number of applicants and limited space. Students who are aware of these programs might apply for upwards of 10-15 positions to maximize the chance of acceptance, and once accepted these programs often require travel which may be a financial hurdle for students. Undergraduate institutions may also have internal programs for underrepresented minorities. For example, Boise State University offers the Bridge to Baccalaureate (B2B, requires full time enrollment), the Higher Education Research Council (HERC) Fellowship (available only to STEM majors), and LSAMP summer research experience (requires students to be of minority ethnic backgrounds). Eligibility for these internal programs focus on ethnic minority groups or require students to be STEM majors. The

B2B program does widen the eligibility to other minority groups (i.e., rural residents, first generation students, disabled students, etc.), but may still exclude other marginalized groups (i.e., women, military, low-income, LGBTQ+, foreign-born). To overcome some of these barriers, Vertically Integrated Projects (VIPs) offer a research ecosystem where undergraduate and graduate students enroll in a course with faculty mentors, where students participate in multidisciplinary research experience in a diverse environment (Strachan et al., 2019). That being said, VIPs may still exclude students who may have financial difficulties and are not able to pay for the extra course credits once the allowable four credits towards a degree program are met.

Expanding research opportunities to a larger and more diverse cohort and integrating multiple disciplines can be achieved by created authentic research experiences that are embedded within the required undergraduate curriculum. These Course-Based Undergraduate Research Experiences (CUREs) and Lab-Based Undergraduate Research Experiences (LUREs) modules within a course offer an emerging opportunity for undergraduate students to receive authentic research experiences as part of their existing degree programs (Peteroy-Kelly et al., 2017). In these courses and labs, students translate content-based knowledge and standard laboratory skills into novel research questions that generate outcomes or predictions that are of interest to outside stakeholders. Simultaneously, students gain real world experience that can benefit multiple career tracks by learning to be self-sufficient, collaborative, and innovative scientists.

CURES and LUREs can also provide an opportunity to integrate knowledge across disciplines. For example, undergraduate students in biology degree programs receive a background in STEM through required biology, ecology, chemistry,

physiology, math, and communication courses in addition to electives. For example, the Bachelor of Science in Biology at Boise State University requires that students take general science and math courses (chemistry, biology, physics, statistics, etc.) along with more advanced courses (organic chemistry, human/animal/plant physiology, calculus, etc., Appendix B). However, students rarely have the opportunity to intentionally integrate concepts and techniques from different STEM fields into one lab (Aikens, 2020). Understanding how various STEM topics and disciplines connect with each other is critical for the future of these students when taking on careers in the biological sciences. CUREs and LUREs can provide students with the opportunity to learn about and practice concepts in a single lab where they intentionally integrate disciplines to discover an unknown pattern or solve a scientific problem. The standard undergraduate curriculum results in students spending hours focusing on single topic each semester, but STEM careers and technological advances require that our future workforce is able to think in interdisciplinary ways. Authentic research experiences that integrate multiple disciplines can help prepare undergraduates to better understand and solve complex scientific problems. While these integrative opportunities are core to the workforce development goals of federal funding agencies (Gross, 2004), they are often not available to large populations of undergraduates.

We created a pharmacokinetic (PK)-based LURE module that provided an authentic research experience that integrated multiple disciplines for a diverse cohort of students enrolled in an Animal Physiology and Nutrition course during the Fall 2021 and Fall 2022 semesters. In general, our PK LURE provided students experiences that focused on scientific practices, discoveries, relevance, collaboration, and was iterative

(Fig 2.1). The purpose of this PK LURE was to integrate basic concepts and scientific practices in chemistry, computing, and physiology that have relevance to practitioners in pharmacology and conservation. Students were also given the opportunity to discover new knowledge by practicing skills in literature searches and using databases PubChem (Kim et al., 2021) and SwissADME (Daina et al., 2017). They collaborated with each other and with a graduate researcher by formulating hypotheses, synthesizing results, and sharing their disciplinary knowledge through scientific writing. At the same time students validated the output of my pipeline using this PK LURE lab module (Chapter 1, Figure 2.2).

Methods

Establishing the Relevance of Pharmacokinetics

Prior to this PK LURE, students were given a lecture-based introduction to PK that focused on how absorption, distribution, metabolism, and excretion (hereafter ADME) play a role in animal physiology (Appendix C). Students were given real world examples through scientific papers with an emphasis on hypotheses that were previously tested within the fields of animal physiology (e.g., (Takahashi & Shimada, 2008)), pharmacology (e.g., (Radeva-llieva et al., 2022)), and conservation of natural resources (e.g., (Parham et al., 1997)). Students were provided these papers as background for part of a current master's student's degree program and to provide relevance across multiple disciplines and emphasized the integration of physiology and chemistry (see Chapter 1). The questions or hypotheses shared with students included:

Chemical Ecology System 1: we examined the following questions using the Japanese wood mouse: (1) does a wood mouse with previous experience eating acorns

select against tannin content and/or select for protein content in acorns more than a wood mouse without previous experience and (2) does previous experience with acorns affect selective consumption? (Takahashi & Shimada, 2008)

Biomedical System 2: What influence do single and multiple oral doses of methylxanthine fraction isolated from *Bancha* green tea leaves have on the pharmacokinetics of sildenafil in rats. (Radeva-llieva et al., 2022)

Toxicology System 3: What are the individual and combined toxicities of different chlorinated PCBs (PCB28, PCB52, and PCB101) on the earthworm *Eisenia fetida* and what is the relationship with the chlorine substitution pattern. (Zhang et al., 2023).

Scientific Practices: Literature Searches

Students were presented with a list of chemicals found in a review paper (Turi et al., 2014) and each student was instructed to conduct literature searches before choosing a chemical that was of interest to them. Allowing students to select a chemical based on their review of the literature provided a degree of ownership and curiosity for students. Using this chemical, each student conducted a preliminary search for peer-reviewed manuscripts to obtain scientific information about the biological significance of the chemical they chose and provided a summary about the information that was found using a properly formatted citation.

Discovery of PK data from a Database

Students were shown an example of how to conduct searches for compounds on PubChem and generate PK information on SwissADME to provide integration of chemistry and computing (Chapter 1). Students used their chosen chemical to obtain

chemical information such as chemical formula, weight, and Canonical Simplified Molecular Input Line Entry System (SMILES) on PubChem, while also getting familiar with a database that they had not used before. Students then navigated to the SwissADME database and used the Canonical SMILES notation they obtained from PubChem to discover PK information (lipophilicity, GI absorption, CYP inhibition, bioavailability, etc.) about their chosen chemical.

Collaboration and Dissemination of Knowledge

Once students had the opportunity to review the ADME parameters of their individual chemicals, they then collaborated with other students in their lab teams sharing and comparing ADME properties to determine which of their chemicals was most likely to have a biological effect in the consumer and what ADME parameters could be used to predict this effect. Once students determined this information, their team came up with a directional hypothesis about how an animal (wild, domestic, or human) would interact with their selected chemical and what the pharmacological outcome would occur through this interaction. They supported their hypothesis using a literature review. Each team identified a representative to orally present a short summary to the rest of class of their rationale for hypothesis, the ADME features of their chemical of interest, and how the reference they found supported or refuted their hypothesis (see Appendix D for examples).

Assessment of Competencies.

Students were given a survey to assess their perceived change in competency in skills learned after the PK LURE which contained questions assessing the knowledge

they gained during the PK LURE module as well as the overall course. (Picture 2.1 and 2.2).

Results

I developed and implemented a novel interdisciplinary LURE module (Appendix E) that will be used each year as a remote authentic research experience for undergraduates. The PK LURE was delivered to 26 students in ZOOL 409 Fall semester of 2022 and was made available to the larger Idaho community through the NSF Idaho EPSCoR website called GEM3 Lab Modules (https://www.idahogem3.org/lab-modules).

Students explored 26 chemicals representing monoterpenes and phenolics (two different chemical classes). The ADME output generated by students was used to validate output from research done for a graduate research thesis project (see Chapter 1). Students identified 7 peer-reviewed papers that covered topics relevant to the class including enzyme kinetics, permeability, and pharmacokinetics. Of these, 100% were new to the MS researcher and helped advance knowledge of her thesis research.

The class generated 7 novel hypothesis and summaries of their rationale for hypothesis, the ADME features of their chemical of interest, and how the reference they found supported or refuted their hypothesis (Appendix D).

Based on the competency assessment survey, we found that after this lab module, 77% of students felt extremely more competent in database navigation and interpretation, 81% of students felt extremely more competent in formulating directional scientific hypotheses, and 85% of students felt extremely more competent in identifying and summarizing literature to support their hypothesis (Figure 2.3). All but one student felt they increased their competency in activities focused on integrating disciplines, using

databases, and testing hypotheses. We also found 96% of students felt extremely or slightly more competent in proposing and testing hypotheses that use chemical knowledge to predict physiological responses in animals, 100% of students were more competent in understanding why chemical structures matter in physiology, and 100% of students were more competent using publicly available databases to conduct research (Figure 2.4).

Discussion

Authentic experiences in database searches and group discussions led to students who were better able to: 1) understand the relevance of pharmacokinetics in animal physiology; 2) practice integration of skills and knowledge in chemistry, computing, physiology, pharmacology, and ecology; and 3) explore how this integration can advance science. Students participating in the PK LURE met lab objectives of conducting searches on publicly available databases and interpreting results (Pubchem and SwissADME). They also strengthened communication skills by comparing and contrasting PK properties in small groups to determine which chemicals were more likely to have biological effect in consumers. This experience allowed them to formulate novel hypotheses to determine outcomes of chemical interactions within the body and conduct literature searches to support their formulated hypothesis within a lab setting. Each of these objectives, combined with the optional remote aspect of this lab, allowed us to reach a larger and more diverse cohort of students while also integrating multiple disciplines into the LURE module.

The PK LURE provided students a unique opportunity to strengthen their Knowledge Skills and Abilities (KSA) in database navigation and interpretation,

formulating directional scientific hypotheses, identifying and summarizing literature to support their hypothesis, proposing and testing hypotheses that use chemical knowledge to predict physiological responses in animals, understanding why chemical structures matter in physiology, and using publicly available databases to conduct research (Figure 2.3, 2.4). Each of these KSAs can be used to help students highlight their knowledge when interviewing for prospective job opportunities (Lopatto, 2007). For example, students interested in disciplines like animal physiology (veterinary medicine/research) (Dale et al., 2010) can speak to the strong communication and research skills associated with how animals interact to chemicals. Students interested in pharmacology (Toews et al., 2005) can speak to PK knowledge and the chemical skills. Students interested in conservation of natural resources (Midden et al., 2007) gained communication skills in terms of public speaking and understanding plant-herbivore interactions. As such the KSA in physiology, chemistry, pharmacology, and computing gained in this LURE created a more competent workforce and made our undergraduates more competitive for veterinary medicine, pharmaceutics, and natural sciences job opportunities.

References

Aikens, M. L. (2020). Meeting the needs of a changing landscape: Advances and challenges in undergraduate biology education. *Bulletin of Mathematical Biology*, *82*(5), 60. https://doi.org/10.1007/s11538-020-00739-6

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., Lawrie, G., McLinn, C. M., Pelaez, N., Rowland, S., Towns, M., Trautmann, N. M., Varma-Nelson, P., Weston, T. J., & Dolan, E. L. (2014). Assessment of course-based undergraduate research experiences: A meeting report. *CBE—Life Sciences Education*, *13*(1), 29–40. https://doi.org/10.1187/cbe.14-01-0004

Beck, C., Butler, A., & Burke da Silva, K. (2014). Promoting inquiry-based teaching in laboratory courses: Are we meeting the grade? *CBE—Life Sciences Education*, *13*(3), 444–452. https://doi.org/10.1187/cbe.13-12-0245

Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. *CBE—Life Sciences Education*, *14*(1), es1. https://doi.org/10.1187/cbe.14-10-0167

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/srep42717

Delgado, T., Bhark, S.-J., & Donahue, J. (2021). Pandemic teaching: Creating and teaching cell biology labs online during COVID-19. *Biochemistry and Molecular Biology Education*, *49*(1), 32–37. https://doi.org/10.1002/bmb.21482

Dustman, W. A., King-Keller, S., & Marquez, R. J. (2021). Development of gamified, interactive, low-cost, flexible virtual microbiology labs that promote higher-order thinking during pandemic instruction. *Journal of Microbiology & Biology Education*, *22*(1), ev22i1.2439. https://doi.org/10.1128/jmbe.v22i1.2439

Estrada, M., Hernandez, P. R., & Schultz, P. W. (2018). A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into STEM careers. *CBE—Life Sciences Education*, *17*(1), ar9. https://doi.org/10.1187/cbe.17-04-0066

Gross, L. J. (2004). Interdisciplinarity and the undergraduate biology curriculum: Finding a balance. *Cell Biology Education*, *3*(2), 85–87. https://doi.org/10.1187/cbe.04-03-0040

Ing, M., Burnette, J. M., Azzam, T., & Wessler, S. R. (2021). Participation in a course-based undergraduate research experience results in higher grades in the companion lecture course. *Educational Researcher*, *50*(4), 205–214. https://doi.org/10.3102/0013189X20968097

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Research*, *49*(D1), D1388–D1395. https://doi.org/10.1093/nar/gkaa971

Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. *CBE—Life Sciences Education*, *6*(4), 297–306. https://doi.org/10.1187/cbe.07-06-0039

Midden, C. J. H., Kaiser, F. G., & Teddy McCalley, L. (2007). Technology's four roles in understanding individuals' conservation of natural resources. *Journal of Social Issues*, *63*(1), 155–174. https://doi.org/10.1111/j.1540-4560.2007.00501.x

Parham, F. M., Kohn, M. C., Matthews, H. B., Derosa, C., & Portier, C. J. (1997). Using structural information to create physiologically based pharmacokinetic models for all polychlorinated biphenyls. *Toxicology and Applied Pharmacology*, *144*(2), 340–347. https://doi.org/10.1006/taap.1997.8139

Peteroy-Kelly, M. A., Marcello, M. R., Crispo, E., Buraei, Z., Strahs, D., Isaacson, M., Jaworski, L., Lopatto, D., & Zuzga, D. (2017). Participation in a year-long CURE embedded into major core genetics and cellular and molecular biology laboratory courses results in gains in foundational biological concepts and experimental design skills by novice undergraduate researchers. *Journal of Microbiology & Biology Education*, *18*(1), 18.1.10. https://doi.org/10.1128/jmbe.v18i1.1226

Radeva-llieva, M., Stoeva, S., Hvarchanova, N., Zhelev, I., & Georgiev, K. D. (2022). Influence of methylxanthines isolated from *Bancha* green tea on the pharmacokinetics of sildenafil in rats. *DARU Journal of Pharmaceutical Sciences*, *30*(1), Article 1. https://doi.org/10.1007/s40199-022-00433-z

Strachan, S. M., Marshall, S., Murray, P., Coyle, E. J., & Sonnenberg-Klein, J. (2019). Using vertically integrated projects to embed research-based education for sustainable development in undergraduate curricula. *International Journal of Sustainability in Higher Education*, *20*(8), 1313–1328. https://doi.org/10.1108/IJSHE-10-2018-0198

Takahashi, A., & Shimada, T. (2008). Selective consumption of acorns by the Japanese wood mouse according to tannin content: A behavioral countermeasure against plant secondary metabolites. *Ecological Research*, *23*(6), Article 6. https://doi.org/10.1007/s11284-008-0473-5

Turi, C. E., Shipley, P. R., & Murch, S. J. (2014). North American *Artemisia* species from the subgenus *tridentatae* (Sagebrush): A phytochemical, botanical and pharmacological review. *Phytochemistry*, *98*, 9–26. https://doi.org/10.1016/j.phytochem.2013.11.016

Zhang, J., He, M., Liu, Y., Zhang, L., Jiang, H., & Lin, D. (2023). Chlorine substitution-dependent toxicities of polychlorinated biphenyls to the earthworm *Eisenia fetida* in soil. *Journal of Environmental Sciences*, *128*, 171–180. https://doi.org/10.1016/j.jes.2022.07.032

Figures

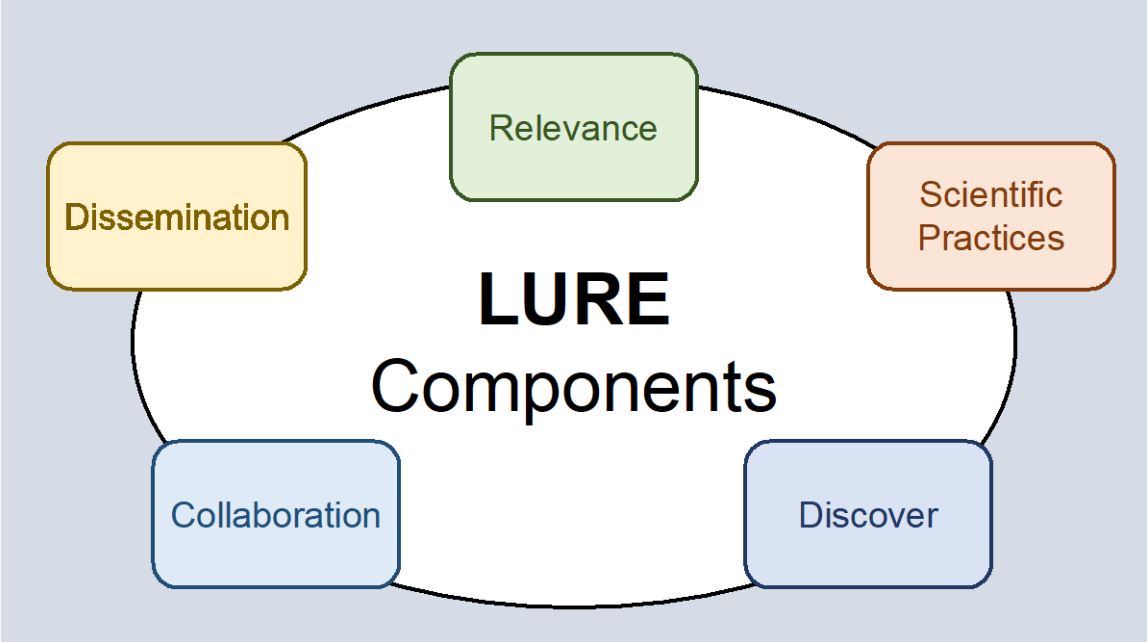


Figure 2.1 The making of a Lab-based Undergraduate Research Experience (LURE) lab. Students used 'Relevant' (remote) 'Scientific Practices' in computing to 'Discover' pharmacokinetic information about compounds that influence animal physiology and 'Collaborated' with each other and a graduate student to 'Iteratively' generate new data, relevant literature, and hypotheses.

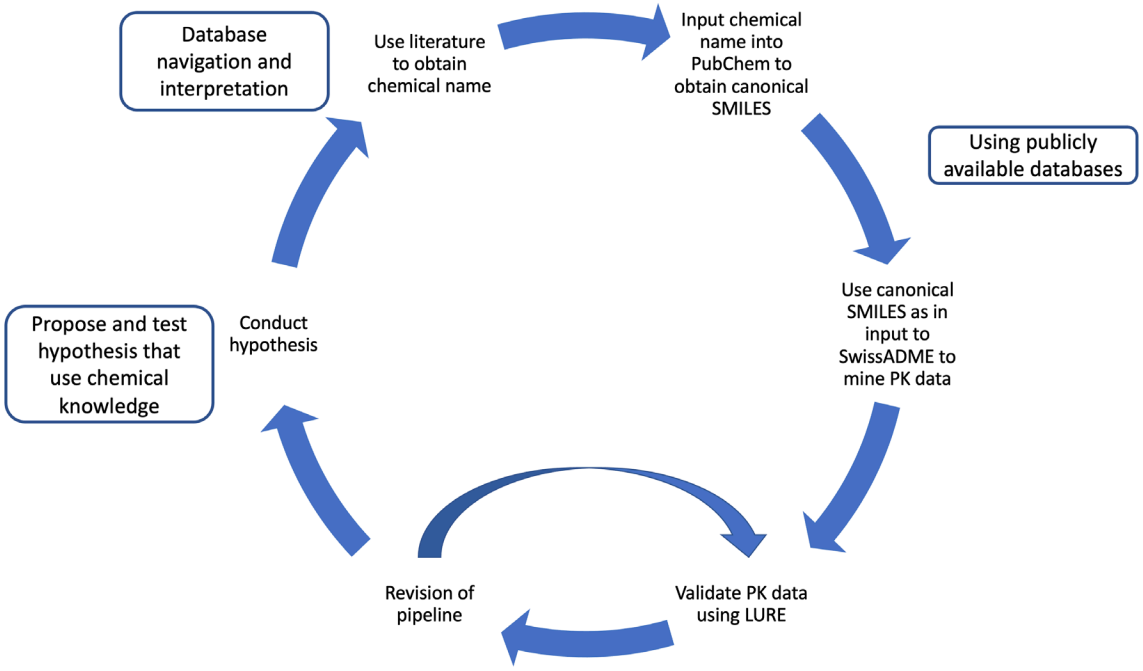


Figure 2.2 The cycle of how the pharmacokinetic (PK) Lab-based Undergraduate Research Experience (LURE) module fits into the validation step of PK output and revision an automated pipeline in Chapter 1. Specific areas of scientific training that increased competency of undergraduate students (Figure 2.3) are shown in outer boxes.

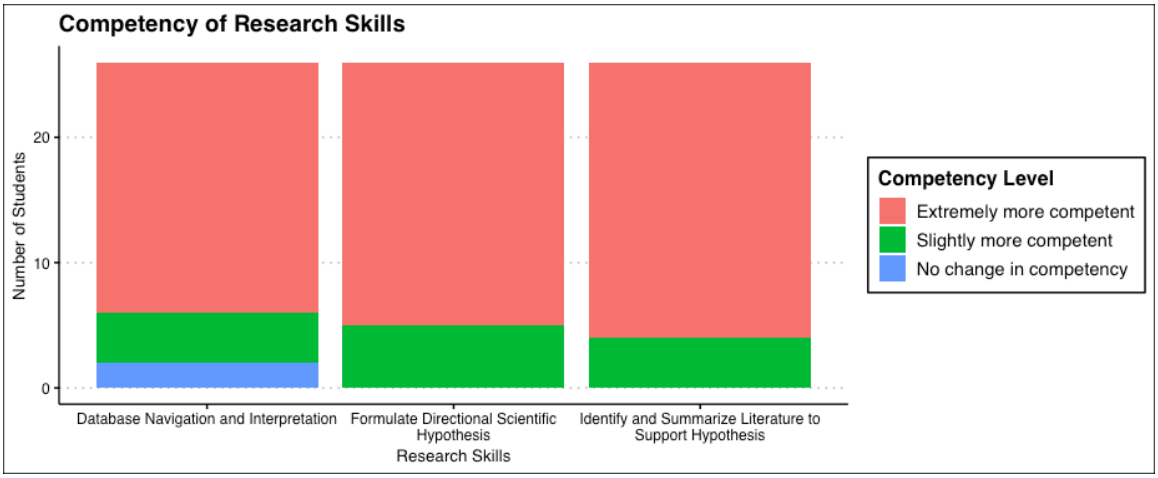


Figure 2.3 Change in competency perceived by undergraduate students in research skills learned and reinforced in the Pharmacokinetic Lab-Based Undergraduate Research Experience module.

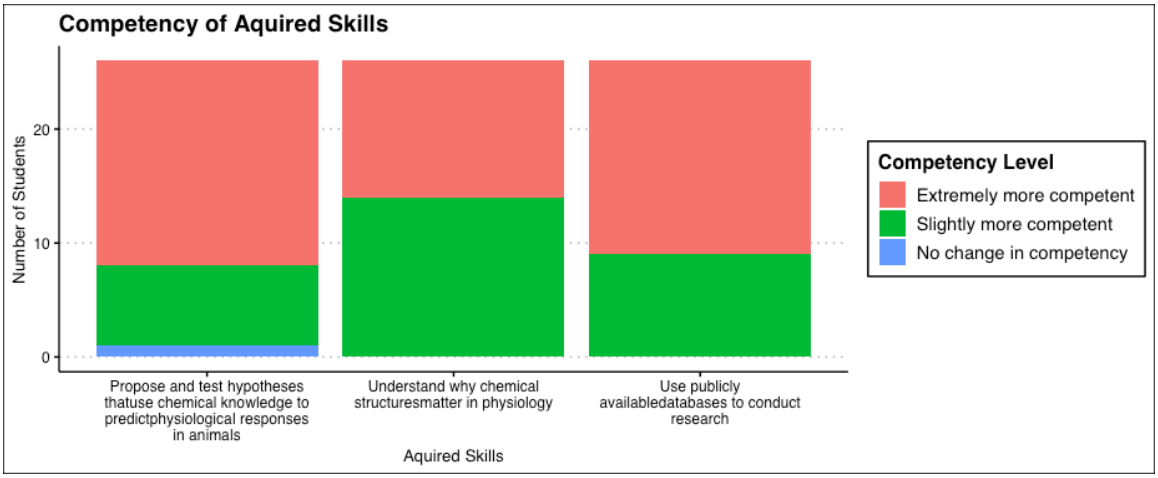


Figure 2.4 Change in competency perceived by undergraduate students in scientific skills learned and reinforced in the Pharmacokinetic Lab-Based Undergraduate Research Experience module.

Pictures

Picture 2.1 Knowledge, Skills, and Abilities (KSA) Survey question on research skills.

ADME lab module: How did the ADME lab module led by Celin change your competency in the following areas?

	Extremely more competent	Slightly more competent	No change in my competency
I understand why chemical structures matter in physiology	○	○	○
I know how to use publicly available databases to conduct research	○	○	○
I could propose and test hypotheses that use chemical knowledge to predict physiological responses in animals	○	○	○

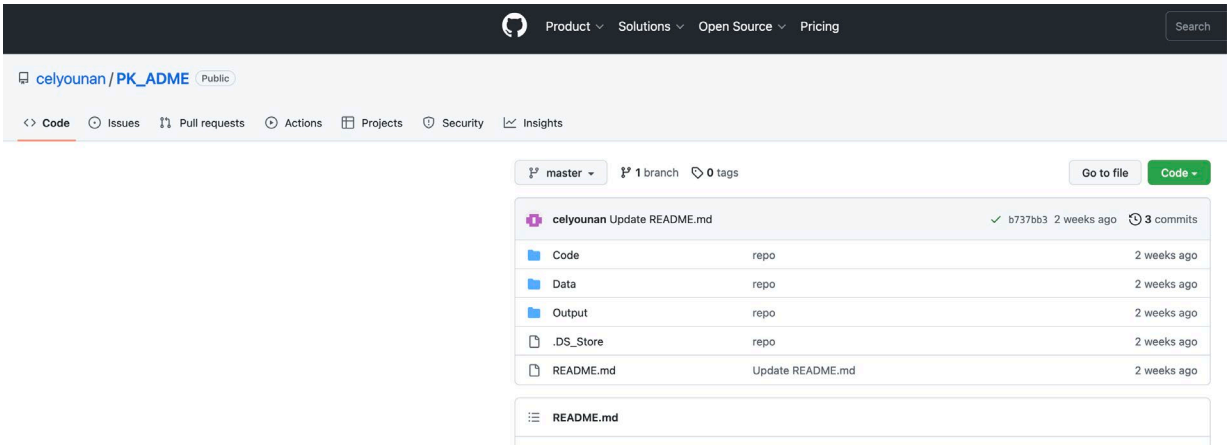
Picture 2.2 Knowledge, Skills, and Abilities (KSA) Survey question about ADME lab outcomes.

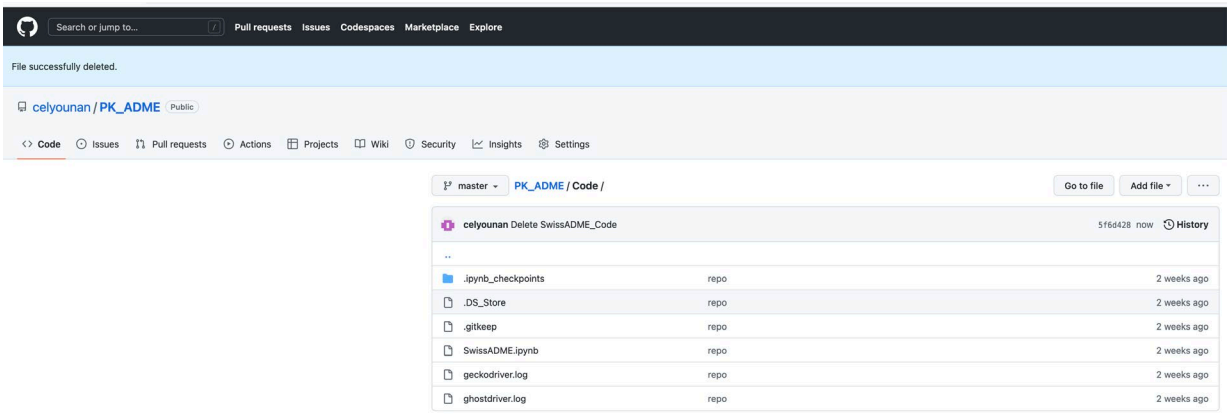
APPENDIX A

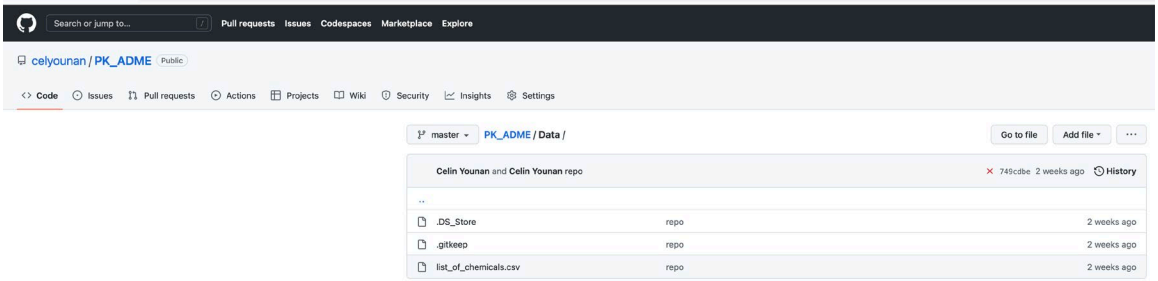
Link to GitHub Repository and Screen Shots of Repository

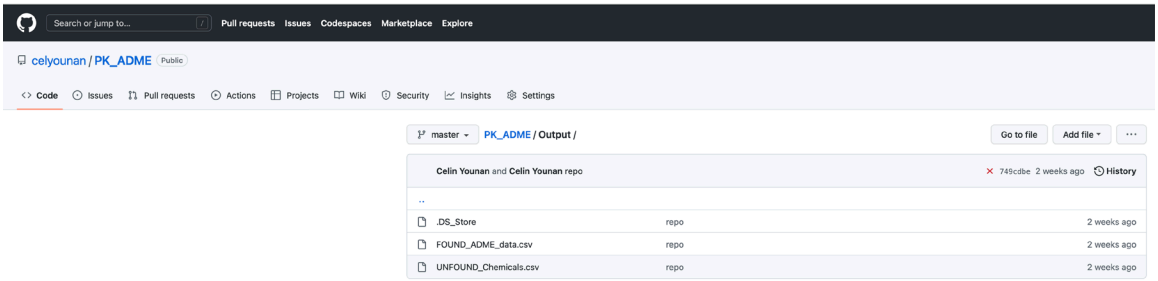
Link and screenshots of GitHub repository showing file structure of the project.

GitHub link: https://github.com/celyounan/PK_ADME






Search or jump to...
Pull requests
Issues
Codespaces
Marketplace
Explore
celyounan / PK_ADME
Public
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
master
PK_ADME / Data /
Go to file
Add file
Celin Younan and Celin Younan repo
749cdbe 2 weeks ago
History
..
.DS_Store
repo
2 weeks ago
.gitkeep
repo
2 weeks ago
list_of_chemicals.csv
repo
2 weeks ago


Search or jump to...
Pull requests
Issues
Codespaces
Marketplace
Explore
celyounan / PK_ADME
Public
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
master
PK_ADME / Output /
Go to file
Add file
Celin Younan and Celin Younan repo
749cdbe 2 weeks ago
History
..
.DS_Store
repo
2 weeks ago
FOUND_ADME_data.csv
repo
2 weeks ago
UNFOUND_Chemicals.csv
repo
2 weeks ago

APPENDIX B

Sample Undergraduate Biology Degree Plan

Sample undergraduate biology degree plan. Possible courses with number of credits. Courses in Red are courses that have content integrated and reinforced in the Pharmacokinetic Lab-based Undergraduate Research Experience module.

Sample Undergraduate Degree Plan (Page 1)

Year One – First Semester

Subject/Course	Title	Credits
Biol 191	Biology 1: Intro to Cell and Molecular Biology	4
Math 254	Introduction to Statistics	3
UF 100	Foundations of Intellectual Life	3
Engl 101	Intro to College Writing	3
Total		**13**

Year One – Second Semester

Subject/Course	Title	Credits
Biol 192	Biology II: Intro to Diversity of Life	4
Math 143	College Algebra I	3
FA/FS	Foundations of Art Or Foundations of Social Sciences	3
Engl 102	Intro to College Writing	3
Comm 101	Fundamentals of Oral Communication	3
Total		**16**

Year 2 – First Semester

Subject/Course	Title	Credits
Biol 304	Biology III: Ecology & Evolution	4
CHEM 111	General Chemistry I	3
CHEM 111L	General Chemistry I Lab	1
MATH 144	College Algebra II	2
UF 200	Foundations of Ethics and Diversity	3
FA/FS	Foundations of Art or Foundations of Social Sciences*	3
Total		**16**

Year 2 – Second Semester

Subject/Course	Title	Credits
BIOL 310	Genetics	3
CHEM 112	General Chemistry II	3
CHEM 112L	General Chemistry II Lab	1
Choose One: MATH 160 MATH 170	Survey of Calculus or Calculus I	4
FS	Foundations of Social Sciences in 2nd field	3
FH	Foundations of Humanities	3
Total		**17**

Sample Undergraduate Degree Plan (Page 2)

Year 3 – First Semester

Subject/Course	Title	Credits
CMB Class w/ Lab	Refer to Catalog	4
BIOL 320	Cell Biology	3
CHEM 307	Organic Chemistry I	3
CHEM 308	Organic Chemistry I Lab	2
Required Communication Course	Refer to Catalog	3
Total		**15**

Year 3 – Second Semester

Subject/Course	Title	Credits
CMB Class w/o Lab	Refer to Catalog	3
BIOL 400	Organic Evolution	3
PHYS 111 or 211	General Physics I or Physics with Calculus I	4 or 5
Elective		3
Required Communication Course	Refer to Catalog	3
Total		**16**

Year 4 – First Semester

Subject/Course	Title	Credits
EEB Class w/ Lab	Refer to Catalog	3
Physiology Requirement	Refer to Catalog	3 or 4
PHYS 112 or 212	General Physics II or Physics with Calculus II	4 or 5
Upper-division Elective		3
Elective		3
Total		**16**

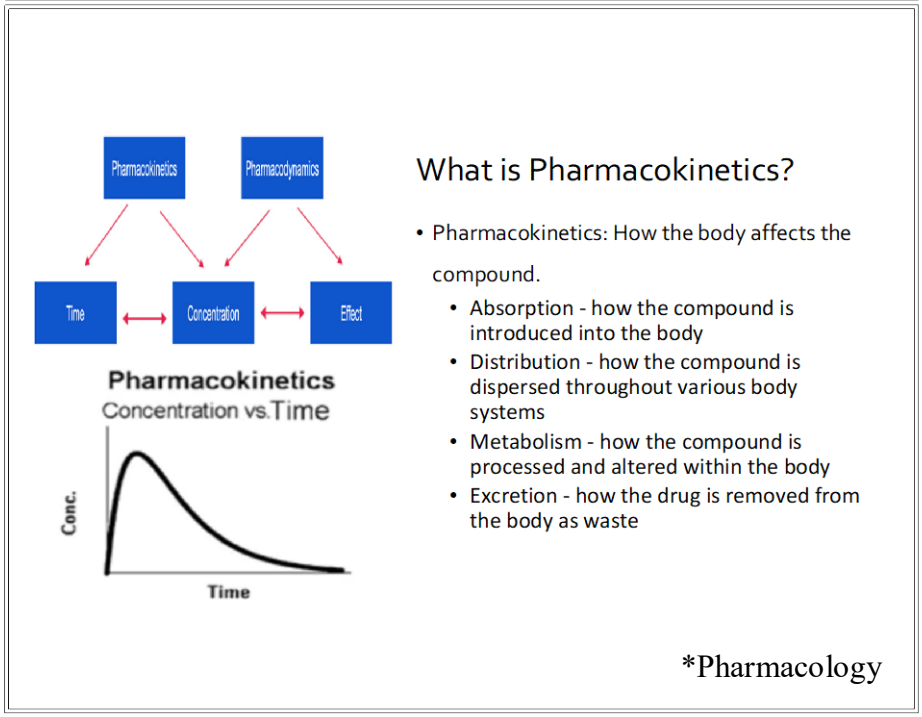
Year 4 – Second Semester

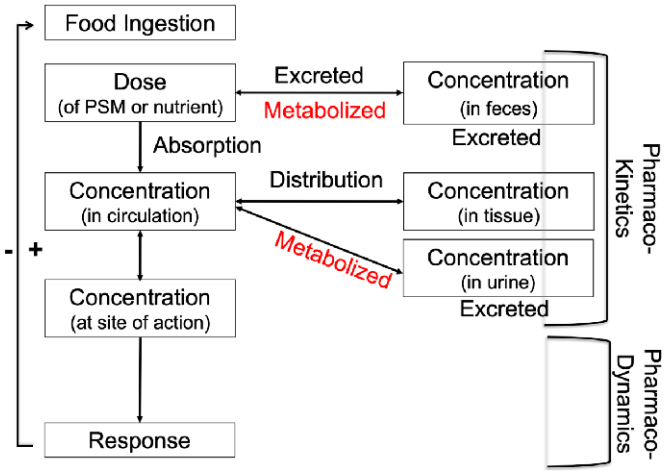
Subject/Course	Title	Credits
EEB Class w/o Lab	Refer to Catalog	3
Finishing Foundation Course	Refer to Catalog - physiology course (animal, human, microbial or plant)	4
Upper-division Elective		3
BIOL 488	Senior Outcomes Assessment	0
Upper Division Elective		3
Total		**13**

APPENDIX C

PowerPoint Presentation Presented to Undergraduate Students in LURE Course

LURE PowerPoint presentation on Pharmacokinetics (ADME) Lab presented to Animal Physiology and Nutrition course where specific topics from their degree program and integration of disciplines were emphasized (as note indicated with an asterisks).

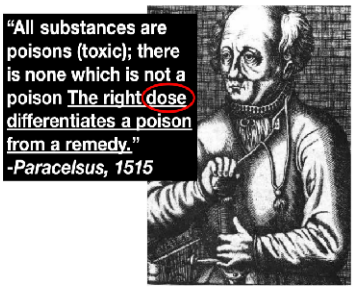




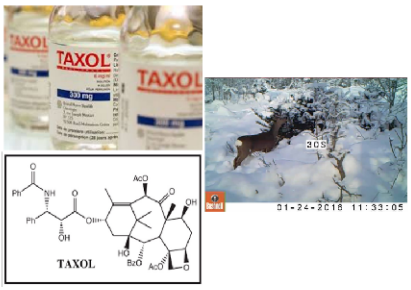
*Physiology

Why PK-PD?

All things are toxic - Paracelsus



Humans want to determine doses and treatments

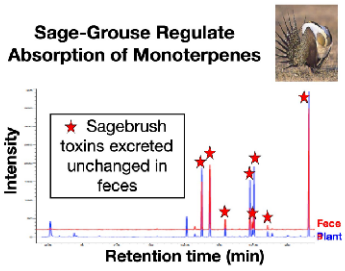


*Pharmacology

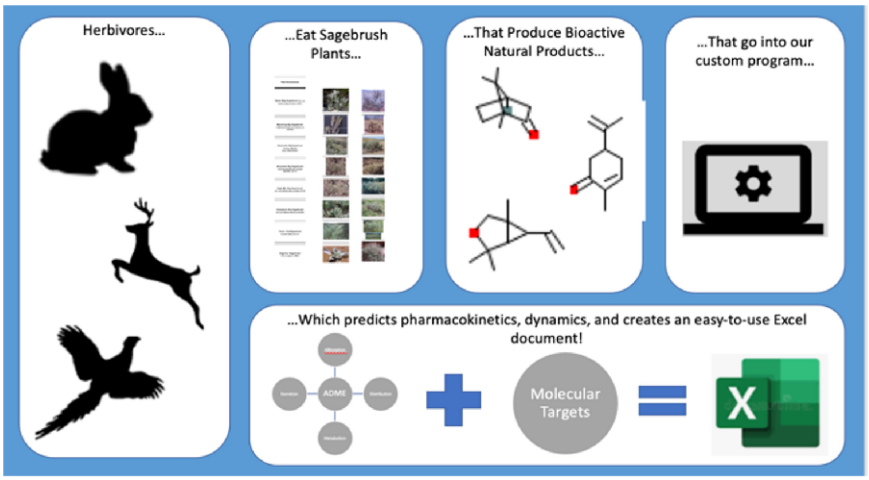
Why PK-PD?

Wildlife rely on detecting and regulating doses of dietary or environmental toxins

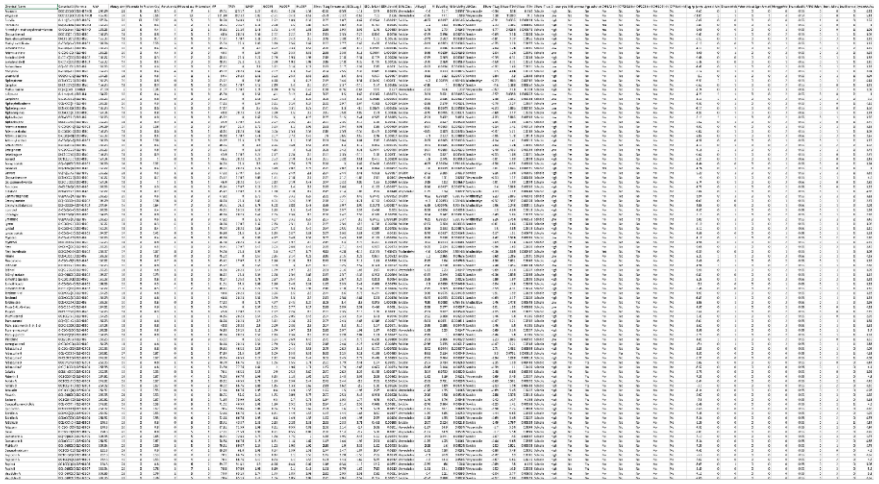
Use concept of PK and PD to predict interactions with chemicals.



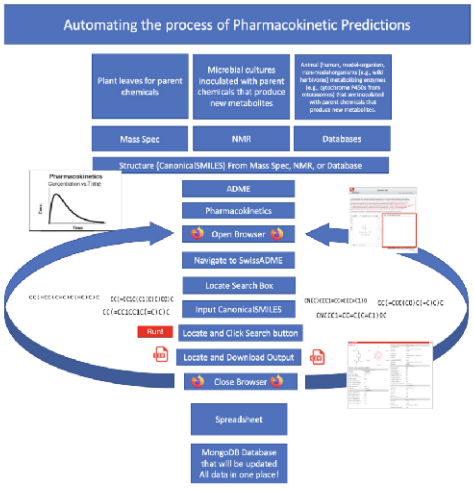
*Ecology



*Chemistry



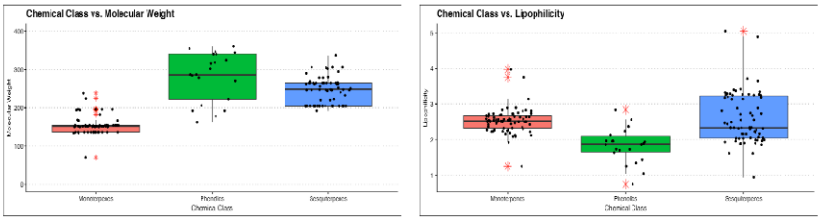
*Computing



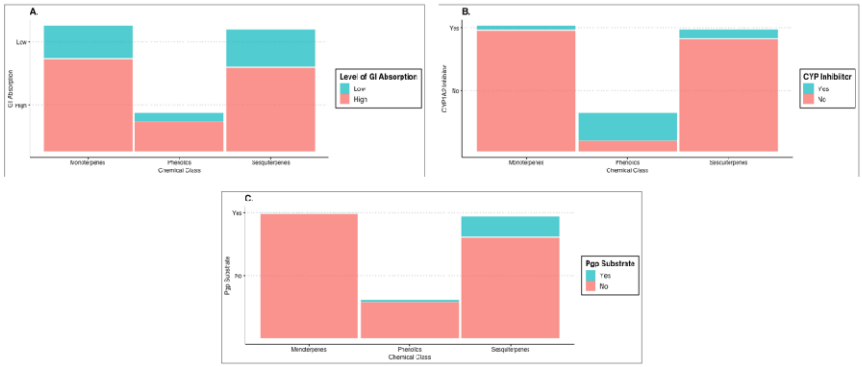
Schematic of the steps taken to automate the process of pharmacokinetic predictions

*Computing

Hypothesis: We first tested the hypothesis that because monoterpenes are smaller in molecular weight and have high lipophilicity, they would have higher GI absorption, are more likely to be CYP inhibitors, and are more likely to be P-gp substrates than phenolics and sesquiterpenes.



*Statistics



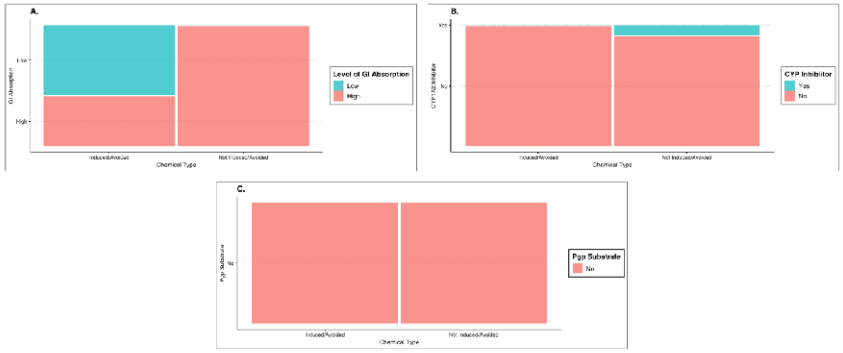
Predicted Gastrointestinal (GI) Absorption (A, high versus low), CYP1A2 Inhibition (B, yes versus no), and Pgp Substrate (C, yes versus no) of monoterpene (n=73), phenolic (n=22), and sesquiterpenes (n=71) identified in sagebrush (*Artemisia* spp.)

*Statistics

Second, we hypothesized that within monoterpenes, those that are induced or avoided by foraging herbivores would have higher GI absorption, are more likely to be CYP inhibitors, and are less likely to be P-gp substates compared to the individual monoterpenes that are not induced or avoided by herbivores

Induced/Avoided	Not Induced/Avoided
Methacroleine	Santolinolide-B
p-Cymene	Lyratyl acetate
Alpha-pinene	Terpinen-4-ol
Camphene	Geranyl acetate
Alpha-phellandrene	Neryl isovalerate
Beta-pinene	Alpha-ionone
Camphor	Dehydroleucodin
Borneol	Arbiglovin
Artemiseole	Santonin
1,8-cineole	Rothin-A
Beta-caryophyllene	Deacetyllaurenobiolide
Germacrene-D	Pygmol

*Ecology and Chemistry



Predicted Gastrointestinal (GI) Absorption (A, high versus low), CYP1A2 Inhibition (B, yes versus no), and Pgp Substrate (C, yes versus no) of monoterpenes in sagebrush (*Artemisia* spp.) that were observed to be induced or avoided (n=12) or not induced or avoided (n=12) by foraging herbivores.

*Ecology and Chemistry

Lastly, we hypothesized that species or subspecies of *Artemisia* that are less palatable based on selective foraging by vertebrate herbivores would have a higher proportion of chemicals with high GI absorption and CYP inhibition and a lower proportion of chemical that are P-gp substrates than more palatable taxonomic groups.

Species or Subspecies of *Artemisia* that are LESS palatable

A. tridentata subsp. Wyomingensis

A. tridentata subsp. Tridentata

species or subspecies of *Artemisia* that are MORE palatable

A. tridentata subsp. vaseyana

Artemisia arbuscula subsp. longiloba

Artemisia nova

Artemisia arbuscula subsp arbuscula

*Ecology

Idaho Statesman

HOME NEWS SPORTS BUSINESS ENTERTAINMENT OPINION OUTDOORS

News > Local News > **Environment**

Sagebrush: A prescription for new medicine?

With garden shears and some help from wildlife, a College of Idaho researcher might turn up some new medicine

BY BILL ROBERTS

BSU researcher's work could lead to new and more effective drugs

BY KATHLEEN TUCK

Boise State University May 8, 2015

*Medicine

Homework

Homework - Step 10: Navigate to the Turi Paper and take a look at the tables of chemicals separated into chemical classes. Navigate to ADME_ChemList worksheet #3 ADME_ChemList and each student is assigned for 5 chemicals per person and specify on the worksheet what species of sagebrush each of your 5 chemicals came from.

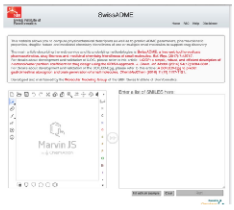
No Short Report this week

*Communication

ADME Lab Module Steps

- Step 1 – Select your chemical
- Step 2 – Pubchem
- Step 3 – SwissADME
- Step 4 – Compare ADME properties with your group
- Step 5 – Identify ADME properties
- Step 6 – Hypothesis
- Step 7 – Lit search to support your hypothesis
- Step 8 – Draft
- Step 9 – Share
- Homework - Step 10 – See Canvas Module XX – ASSIGNMENT NAME NO Short Report this week.

PubChem



*Computing

APPENDIX D

Hypothesis and Summaries for Each Team in LURE Course

Hypothesis and summaries Animal Physiology and Nutrition (LURE course)

students developed in teams.

Team 1

Hypothesis - “CYP1A2 metabolizes a drug used in the treatment of asthma. We hypothesize that someone taking both asthma medication and scoparone would see a reduction in the function of their asthma medication, resulting in more severe asthma symptoms.”

Summary - “The researchers compared the metabolism rate of Theophylline in two groups of rats; one having normal counts of CYP1a2 and one with low counts of CYP1a2 and found that the rats with higher CYP1a2 levels were able to metabolize more Theophylline. This confirms the role of CYP1a2 as a metabolizing agent of theophylline. Theophylline is used in the treatment of respiratory ailments, including asthma.”

Team 2

Hypothesis - “The absorption of prunasin will reduce glucose absorption in mammals.”

Summary - “While testing to find the transport mechanism for prunasin this study found that prusasin when consumed by rats reduces the absorption of glucose in the intestine.”

Team 3

Hypothesis - “We hypothesize that if a human consumed lavandulol, it would interact with their nervous system and have a sedative affect.”

Summary - “This study found that lavadulol (and other chemicals found in lavender oil) have the ability to interact with the nervous system and produce a sedative, mood-stabilizing, and even anti-convulsive affect.”

Team 4

Hypothesis - “Habitual smokers (human) who inhale diffused eucalyptol oil (via diffuser device) alongside experience less severe effects of smoking on the lungs.”

Summary - “Kennedy et al., tested six mice who were exposed to short-term cigarette smoke and a group that was not. The results suggest a reduction in immune response (number of leukocytes and cytokines) and a reduction in reactive oxygen species in mice who smoked but were also exposed to eucalyptol.”

Team 5

Hypothesis - “A human that ingests with this chemical would experience lower stress and anxiety.”

Summary - “Kaempferol increases survival rate in mice when receiving total body gamma irridation because it inhibits oxidative stress. Kaempferol protects tissues and prevents changes in cell morphology that would result in cell apoptosis.”

Team 6

Hypothesis - “We hypothesize that Hordenine's high GI absorption capability, as well as its ability to penetrate the blood brain barrier, could have significant, positive effects on the human brain.”

Summary - “The paper supports our hypothesis that hordenine has positive impacts on human brain fuction. This study shows GI absorption of hordenine was inefficient, but dosing through nasal or intravenous methods results in better absorption and higher brain uptake of the chemical, which overall leads to stronger interactions with dopamine receptors. Hordenine affects dopamine receptors in the brain, which could have positive impacts on mood, and can also be used to help with central nervous system disorders.”

Team 7

Hypothesis - “We hypothesize that borneol's ability to penetrate the blood brain barrier can have beneficial effects on the nervous system in humans.”

Summary - “Researchers wanted to understand borneols ability to be used as a glioma tumor treatment. They tested the abililty of borneol to open the blood brain barrier as a treatment to test its effect on glioma cells in rats. They also tested how borneol influenced apoptosis in cultured human glioma cells, and found that it increased apoptosis of these cells, supporting the hypothesis that borneol can have beneficial influences on human health.”

APPENDIX E

Pharmacokinetics Absorption, Distribution, Metabolism, and Excretion (ADME)

Lab Protocol

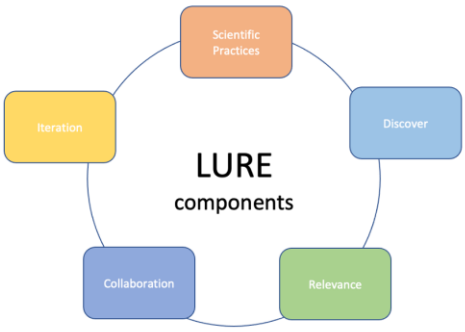
Students were given this pharmacokinetics (ADME) Lab protocol to complete after being presented with a pharmacokinetics presentation (Appendix C).

Title: Pharmacokinetics (ADME) Lab

Objectives

The purpose of this lab is to:

1) Students will be able to conduct searches on publicly available databases and interpret results (Pubchem and SwissADME).
2) Students will be able to compare and contrast chemical ADME properties in small groups to determine which chemical is more likely to have a biological effect in the consumer.



3) Students will formulate hypotheses to determine outcomes of chemical interactions within the body.
4) Students will conduct literature searches to support their formulated hypothesis.

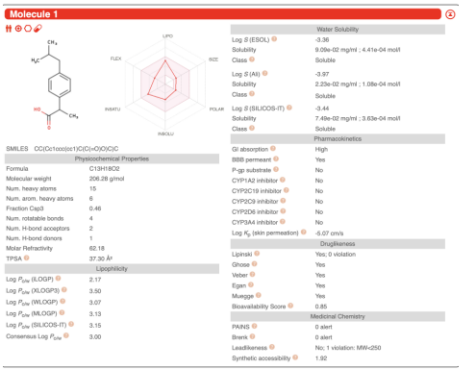
Lab Procedures

Step 1: Each student will select a chemical from "List of chemicals signup" worksheet in: List of Chemicals sign up sheet and be sure that chemical also exists on worksheet #3 "ADME_ChemList" and sign up for that chemical

Step 2: Go to PubChem: https://pubchem.ncbi.nlm.nih.gov/ and get chemical information and enter it in "Swiss ADME info selected chemical hypothesis" (yellow highlighted column) -worksheet in: Swiss ADME info selected chemical hypthesis worksheet 2

Step 3: Go to Swiss ADME site: http://www.swissadme.ch/ and

a. Navigate to "Enter a list of SMILES here:" and in the text box enter your CanonicalSMILES (2.1.4 from pubchem)
b. Click Run
c. Compare your ADME to ADME information found here: ADME_ChemList worksheet #3 ADME_ChemList



Step 4: As a group share and compare ADME properties and decide “which of your group’s chemicals is most likely to have a biological effect in the consumer” enter that info in “Swiss ADME info selected chemical hypothesis” (yellow highlighted column) -worksheet in:
Swiss ADME info selected chemical hypthesis Worksheet 2

Step 5: Identify which ADME parameter(s) is/are used to predict that your selected chemical would have a biological effect on the consumer?

Step 6: Come up with a hypothesis (make sure directional) about how an animal (wild, domestic, or human) would interact with selected chemical and what the outcome would be.

Step 7: Identify literature to support hypothesis and enter Author. year. Title. journal. DOI of that manuscript.

Step 8: Draft a short (no more than 2 sentences) summary of how the reference you found supports or refutes your hypothesis.

Step 9: Select a representative from your group to share ADME features of your chemical of interest and describe rationale for your hypothesis with the rest of class.

Homework - Step 10: Navigate to the Turi Paper and take a look at the tables of chemicals separated into chemical classes. Navigate to ADME_ChemList worksheet #3 ADME_ChemList and sign up for 5 chemicals per person and specify on the worksheet what species of sagebrush each of your 5 chemicals came from.

Helpful Links
SwissADME Paper: https://pubmed.ncbi.nlm.nih.gov/28256516/